Fundamental properties of random Hermitian matrices

Olli Hirviniemi

February 26, 2017

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta/Osasto — Fakultet/Sektion — Faculty	Laitos — Institution — Department
Matemaattis-luonnontieteellinen	Matematiikan ja tilastotieteen laitos

Tekijä — Författare — Author

Olli Hirviniemi

Työn nimi — Arbetets titel — Title

Fundamental properties of random Hermitian matrices

Oppiaine — Läroämne — Subject

Matematiikka

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
Pro gradu -tutkielma	Helmikuu 2017	48 s

Tiivistelmä — Referat — Abstract

In this thesis we examine the properties of Wigner matrices. We will give proofs for two fundamental limit theorems of random Hermitian matrices. One of them is Wigner's semicircular law which states that the distribution of the eigenvalues approaches the Wigner's semicircular distribution when the size of the matrix increases. The other, Bai-Yin theorem, tells that the operator norm of such $n \times n$ matrix is almost surely $(2 + o(1))\sqrt{n}$.

In Chapter 1 we begin by introducing Wigner matrices and deduce the proper normalizing factor. We will also consider what does it mean for the distribution of the eigenvalues to converge by demanding that the empirical spectral distribution should converge almost surely in weak topology.

In Chapter 2 we introduce the Stieltjes transform, a useful tool for finding the limit measure. We prove some basic properties and prove that the weak convergence of measures is equivalent to the convergence of their Stieltjes transforms.

In Chapter 3 we prove the Wigner's semicircular law. The proof is based on using the Stieltjes transform, and contains several steps. First we find the pointwise expectation of the Stieltjes transform by deriving a polynomial equation for it. While deriving this equation we use several techniques and theorems from both probability theory and linear algebra, covered in the appropriate Appendix. After deducing that the expectation converges, we see quite straightforwardly that the pointwise limit is in fact almost sure. From this we deduce that the empirical spectral distribution of the Wigner matrix converges to the Wigner's semicircular distribution.

We begin Chapter 4 by concluding the lower bound in Bai-Yin theorem directly by the semicircular law. For upper bound, we can split the matrix into three parts and use triangle inequality. The diagonal part is easily seen to grow slower than $\sqrt{n}$. The part with smaller elements is estimated using even moments and combinatorics to see that it has the wanted upper bound. Finally the part containing all large elements is almost surely sparse, and therefore its operator norm grows slower than $\sqrt{n}$.

Avainsanat — Nyckelord — Keywords

Random matrices, semicircular law, operator norm, Bai-Yin theorem, limit theorems

Säilytyspaikka — Förvaringsställe — Where deposited

Kumpulan kampuskirjasto

Muita tietoja — Övriga uppgifter — Additional information

Contents

1 Wigner matrices **2**

2 Stieltjes transform **9**

3 Deriving the semicircular law **14**

3.1 Concentration of the quadratic form $C_n^* A_n C_n$ 15

3.2 Finishing the computation 23

4 Operator norm of Wigner matrices **29**

A Some theorems from probability theory **39**

B Hermitian matrices **42**

C Formulas for the elements of matrix inverse **45**

Chapter 1

Wigner matrices

A random matrix is a matrix-valued random variable. The idea of a random matrix dates back to the first half of the 20th century. Wishart [2] used random matrices in statistics and von Neumann [5] studied random matrices to estimate numerical errors.

It was Eugene Wigner [3] who noticed the connection between eigenvalues of symmetric random matrices and spectrum of a heavy atom nucleus. He laid some of the first mathematical foundations of the theory [4] and proved a result that would be known as Wigner's semicircular law. This was a major breakthrough and motivation for the study of random matrices.

Later applications of random matrix theory include telephone encryption and many topics in theoretical physics such as quantum chaos and quantum gravity. [1]. The eigenvalues of random matrices also appear to be connected to the zeroes of the Riemann zeta function. [11] The random matrices have many nice and interesting properties that not only answer the questions posed by applications but are also beautiful on their own.

We will focus on Hermitian matrices. We will prove that under some slight assumptions their eigenvalues are distributed according to Wigner's semicircular law. This will in turn give a lower bound for the operator norm, and we will prove that said bound is tight.

Definition 1.1. Suppose that $(\xi_{ij})_{i,j\in\mathbb{Z}_+}$ are random variables such that $(\xi_{ij})_{i\leq j}$ are independent, $(\xi_{ii})_{i\in\mathbb{Z}_+}$ are identically distributed real random variables, $(\xi_{ij})_{i<j}$ are identically distributed complex random variables, and $\xi_{ij} = \overline{\xi_{ji}}$ for all $i > j$. With such random variables, an associated *Wigner matrix* of size n is the Hermitian matrix

$$W_n = (\xi_{ij})_{1\leq i,j\leq n}.$$

Every ξ_{ij} is assumed to have zero mean and unit variance. We will also assume that the fourth moment of the off-diagonal elements is bounded by some constant K: this is true in many important special cases such as Gaussian unitary ensemble (GUE), where the off-diagonal elements are normally distributed complex numbers and the diagonal elements are normally distributed real numbers. Since W_n is Hermitian, it has n real eigenvalues by spectral theorem (Theorem B.1). Trying to understand how these eigenvalues are distributed is the purpose of this section: namely, deriving the *semicircular law*.

First consider the magnitude of the eigenvalues. We first look at a simple example of $n \times n$ -matrix W_n consisting of ones. If we denote the standard basis of $\mathbb{C}^n$ by $\{e_1, \dots, e_n\}$, we see that for $v = \sum_{j=1}^n e_j$ we have $M_n v = nv$. This shows that even with bounded coefficients, the eigenvalues can grow in magnitude as the size of the matrix increases. So we need to normalize the matrix somehow. The following lemma motivates the normalizing factor we use.

Lemma 1.2. *Let x be a unit vector of $\mathbb{C}^n$. Then for any $\lambda > 0$,*

$$P(|W_n x| > \sqrt{n}\lambda) \leq \frac{8}{\lambda^2}.$$

Proof. Fix a unit vector x. Split $W_n = M_1 + M_2$, where M_1 consist of upper triangular elements of W_n (so M_1 is zero below the main diagonal). If we denote the components of x with x_i and rows of M_1 by R_i, then we get following upper bound using independence and zero mean hypothesis:

$$\begin{aligned} E(|M_1 x|^2) &= E\left(\sum_{i=1}^n |R_i \cdot x|^2\right) = \sum_{i=1}^n \sum_{j=i}^n \sum_{k=i}^n E(\xi_{ij} x_j \overline{\xi_{ik} x_k}) \\ &= \sum_{i=1}^n \sum_{j=i}^n |x_j|^2 E(|\xi_{ij}|^2) = \sum_{i=1}^n \sum_{j=i}^n |x_j|^2 \leq \sum_{i=1}^n \sum_{j=1}^n |x_j|^2 = n \end{aligned}$$

By Markov's inequality,

$$P(|M_1 x|^2 \geq \lambda^2) \leq \frac{n}{\lambda^2},$$

or equivalently

$$P(|M_1 x| \geq \sqrt{n}\lambda) \leq \frac{1}{\lambda^2}.$$

Same argument also applies for M_2 with summing j and k from 1 to $i-1$, and we have

$$P(|M_2 x| \geq \sqrt{n}\lambda) \leq \frac{1}{\lambda^2}.$$

Using the triangle inequality, we can estimate

$$\begin{aligned} P(|W_n x| \geq \sqrt{n}\lambda) &\leq P(|M_1 x| + |M_2 x| \geq \sqrt{n}\lambda) \\ &\leq P\left(|M_1 x| \geq \sqrt{n}\frac{\lambda}{2}\right) + P\left(|M_2 x| \geq \sqrt{n}\frac{\lambda}{2}\right) \leq \frac{8}{\lambda^2}. \end{aligned}$$

□

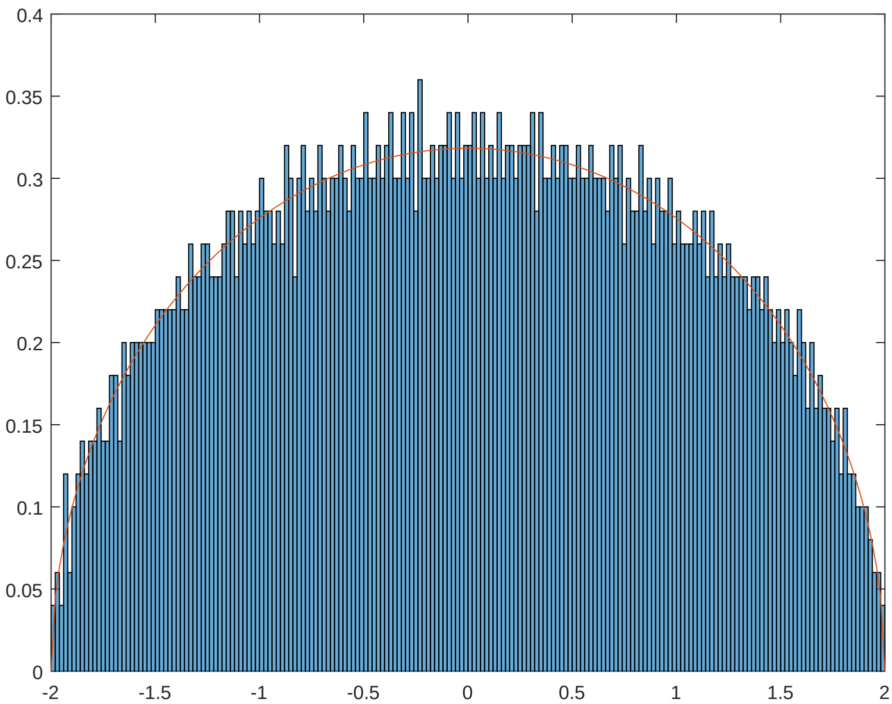

Figure 1.1: Histogram of the normalized eigenvalues from one realization of 2500×2500 random Wigner matrix with off-diagonal elements drawn from standard complex normal distribution and diagonal elements drawn from standard real normal distribution.

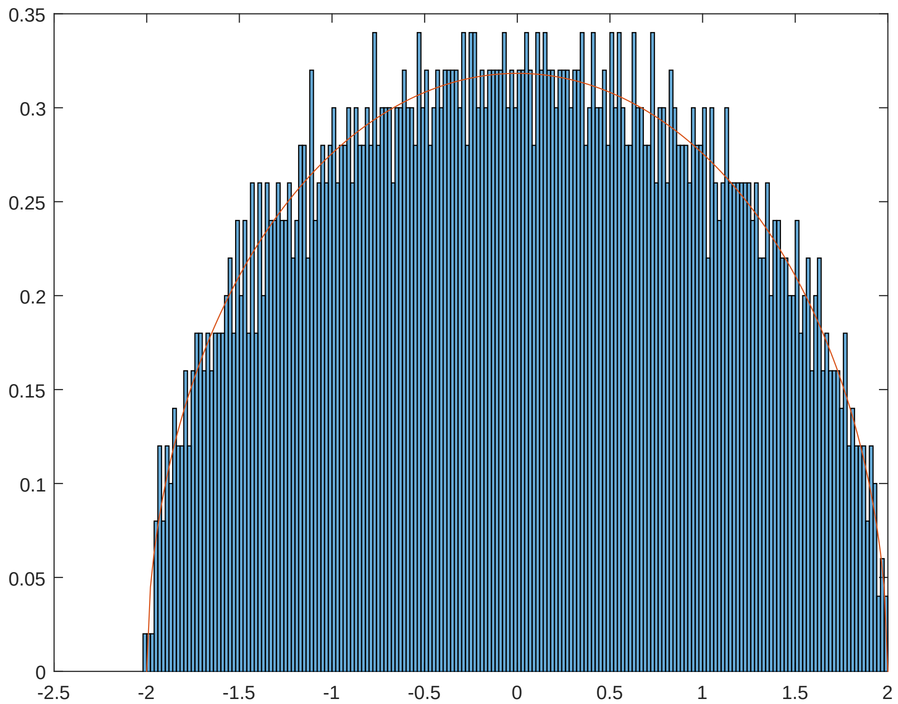

Figure 1.2: Histogram of the normalized eigenvalues from one realization of 2500×2500 random Wigner matrix with elements drawn from standard real normal distribution.

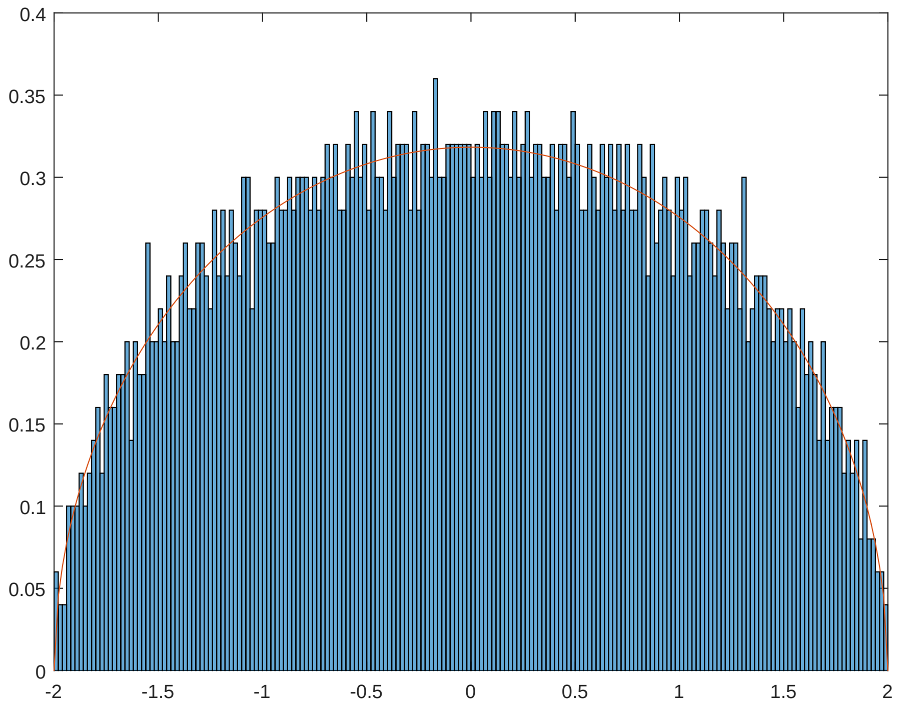

Figure 1.3: Histogram of the normalized eigenvalues from one realization of 2500×2500 random Wigner matrix with elements drawn uniformly randomly from $\{-1, 1\}$.

The previous lemma suggests that we should consider the matrices $\frac{1}{\sqrt{n}}W_n$. After running some numerical simulations (see Figures 1.1, 1.2 and 1.3) it seems likely that the eigenvalues should be distributed approximately according to the *Wigner semicircular distribution* with density

$$\frac{1}{2\pi}\sqrt{\max(0, 4 - x^2)}.$$

To handle this rigorously a precise definition for convergence is necessary.

Definition 1.3. Given a hermitian matrix A_n with eigenvalues $\lambda_1, \ldots, \lambda_n$, the empirical spectral distribution of A_n is a probability measure given by

$$\mu_{A_n} := \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i},$$

where δ_x is a Dirac measure for $x \in \mathbb{R}$.

Denote the space of continuous functions on the real line that vanish at infinity by $C_0(\mathbb{R})$. Recall that a sequence of finite Borel measures μ_n is said to converge to μ weakly if for every $\phi \in C_0(\mathbb{R})$ $\int_{\mathbb{R}} \phi \, d\mu_n$ converges to $\int_{\mathbb{R}} \phi \, d\mu$ as $n \to \infty$.

Example 1.4. For any $y > 0$, we define a function $f_y : \mathbb{R} \to \mathbb{R}_+$ by setting

$$f_y(t) = \frac{1}{\pi}\frac{y}{t^2 + y^2}.$$

Let μ_y be the absolutely continuous measure defined by f_y, i.e. $\mu_y(E) := \int_E f_y(t)\, dt$ for all Borel sets E. Then $\mu_y \to \delta_0$ weakly as $y \to 0$. To see this, let $\phi \in C_0(\mathbb{R})$ be fixed, and make a change of variables to see that

$$\int_{\mathbb{R}} \phi(t) f_y(t)\, dt = \frac{1}{\pi}\int_{\mathbb{R}} \frac{\phi(yx)}{1 + x^2}\, dx.$$

Now the claim follows by the continuity and boundedness of ϕ.

A similar calculation shows that the convolution $\phi * f_y$ converges uniformly to ϕ since $\phi \in C_0(\mathbb{R})$ is uniformly continuous. Integral kernels with this property, such as f_y, are called an approximation of identity.

When A_n is a random hermitian matrix, μ_{A_n} is a random probability measure. We can define the convergence of random measures analogously to the convergence of random variables.

Definition 1.5. Let $(\mu_n)_{n=1}^{\infty}$ be a sequence of random measures. If for any $\phi \in C_0(\mathbb{R})$, the sequence $\int_{\mathbb{R}} \phi \, d\mu_n$ converges to $\int_{\mathbb{R}} \phi \, d\mu$ almost surely, the we say that μ_n converges to μ almost surely.

We can also analogously define the convergence in probability by requiring that for any $\phi \in C_0(\mathbb{R})$, the sequence $\int_{\mathbb{R}} \phi \, d\mu_n$ converges to $\int_{\mathbb{R}} \phi \, d\mu$ in probability.

Our aim is to prove the semicircular law for Wigner matrices, namely that almost surely

$$\mu_{\frac{1}{\sqrt{n}} W_n} \to \mu_{sc} := \frac{1}{2\pi} \sqrt{\max(0, 4 - x^2)}$$

as $n \to \infty$. The outline of the proof is by Tao [10] with some modifications.

Chapter 2

Stieltjes transform

There are different ways to prove the semicircular law. We are going to use the Stieltjes transform for our proof. A different approach would be considering the moments of $\mu_{\frac{1}{\sqrt{n}}W_n}$. This other idea is based on applying the estimates we do later on in Chapter 4 and we will return to it later on.

Definition 2.1. For a finite Borel measure μ on real line, define its *Stieltjes transform* $S\mu$ to be a complex function defined in upper half plane H_+ (or, more generally, outside the support of μ) as

$$S\mu(z) := \int_{\mathbb{R}} \frac{1}{t - z}\, d\mu(t).$$

We will first list some basic properties of the Stieltjes transform.

Lemma 2.2. *For any finite Borel measure μ with $\mu(\mathbb{R}) = C < \infty$, the Stieltjes transform has following properties at any point $z \in H_+$:*

$$\operatorname{Im}(S\mu(z)) > 0 \tag{2.3}$$

$$|S\mu(z)| \leq \frac{C}{\operatorname{Im}(z)} \tag{2.4}$$

$S\mu$ is analytic in H_+ and

$$|S\mu'(z)| \leq \frac{C}{(\operatorname{Im}(z))^2} \tag{2.5}$$

Proof. To prove 2.3, we observe that for any $z = x + iy \in H_+$ and $t \in \mathbb{R}$ we have

$$\operatorname{Im} \frac{1}{t-z} = \operatorname{Im} \frac{(t-x)+iy}{(t-x)^2+y^2} = \frac{y}{(t-x)^2+y^2} > 0.$$

We immediately obtain 2.3 by integrating. Next, we can estimate

$$|S\mu(z)| = \left| \int_{\mathbb{R}} \frac{1}{t-z} \, d\mu(t) \right| \leq \int_{\mathbb{R}} \frac{1}{|t-z|} \, d\mu(t) \leq \int_{\mathbb{R}} \frac{1}{y} \, d\mu(t) = \frac{C}{\operatorname{Im}(z)},$$

obtaining 2.4. It remains to show 2.5. For any fixed $t \in \mathbb{R}$, the function $z \to (t-z)^{-1}$ is analytic in upper half-plane. As the upper half-plane is simply connected, this means that for any closed piecewise smooth curve γ in H_+ we have

$$\int_\gamma \frac{1}{t-z} \, dz = 0.$$

Using Fubini's theorem, we get

$$\int_\gamma S\mu(z) \, dz = \int_\gamma \int_{\mathbb{R}} \frac{1}{t-z} \, d\mu(t) \, dz = \int_{\mathbb{R}} \int_\gamma \frac{1}{t-z} \, dz \, d\mu(t) = 0,$$

so $S\mu$ is analytic by Moreira's theorem. Using Fubini's theorem and Cauchy's integral formula, we have for sufficiently small r

$$S\mu'(z) = \frac{1}{2\pi i} \int_{\partial B(z,r)} \int_{\mathbb{R}} \frac{\frac{1}{t-z}}{(w-z)^2} \, d\mu(t) \, dw = \int_{\mathbb{R}} \left(\frac{d}{dz} \frac{1}{t-z} \right) d\mu(t) = \int_{\mathbb{R}} \frac{1}{(t-z)^2} \, d\mu(t).$$

Using the estimate $|t-z| \geq \operatorname{Im}(z)$ gives 2.5.

□

The following theorem helps us to derive a probability measure if we know its Stieltjes transform.

Theorem 2.6. *Let μ be a finite Borel measure on the real line, and $g_y : \mathbb{R} \to \mathbb{R}$ be the imaginary part of the function $x \to \frac{1}{\pi} S\mu(x+iy)$. For any $\phi \in C_0(\mathbb{R})$, we have*

$$\lim_{y \to 0} \int_{\mathbb{R}} \phi(t) g_y(t) \, dt = \int_{\mathbb{R}} \phi(t) \, d\mu(t).$$

Proof. Let $\phi \in C_0(\mathbb{R})$ be fixed. For $z = x + iy$ define the function $f_y : \mathbb{R} \to \mathbb{R}$ as in 1.4 by setting

$$f_y(t) := \frac{1}{\pi} \frac{y}{t^2+y^2}.$$

Then we see that we can write g_y as convolution of μ and f_y, namely

$$g_y(t) = \int_{\mathbb{R}} f_y(x-t)\, d\mu(t)$$

This means that

$$\lim_{y\to 0}\int_{\mathbb{R}} \phi(t) g_y(t) dt \quad = \quad \lim_{y\to 0}\int_{\mathbb{R}}\int_{\mathbb{R}} f_y(u-t)\phi(t)\, dt\, d\mu(u) = \int_{\mathbb{R}} \phi(u)\, d\mu(u).$$

The last equality holds because $\phi * f_y \to \phi$ uniformly as noted in 1.4. □

Suppose that we only know the Stieltjes transform S_μ of some probability measure μ. If a limit $h(x) := \lim_{y\to 0} \frac{1}{\pi}\operatorname{Im}(S\mu(x+iy))$ exists for almost all $x \in \mathbb{R}$, then we could easily check if the absolutely continuous measure μ_h defined by h would have the correct Stieltjes transform. The following theorem implies that the Stieltjes transform is bijective, so $\mu = \mu_h$.

Theorem 2.7. (Stieltjes continuity theorem, deterministic version.) *Let $(\mu_n)_{n\in Z_+}$ be a sequence of finite Radon measures on the real line and μ be a finite Radon measure on the real line. Assume also that μ_n are uniformly bounded, i.e. $\mu_n(\mathbb{R}) \le C < \infty$ for some constant C. Then $\mu_n \to \mu$ weakly if and only if $S\mu_n(z) \to S\mu(z)$ for all $z \in H_+$.*

Proof. We first assume that $\mu_n \to \mu$ weakly. Then for any $z \in H_+$ the function $t \mapsto (t-z)^{-1}$ is in $C_0(\mathbb{R})$, so by definition of weak convergence

$$S\mu_n(z) = \int_{\mathbb{R}} \frac{1}{t-z}\, d\mu_n(t) \to \int_{\mathbb{R}} \frac{1}{t-z}\, d\mu(t) = S\mu(z).$$

For the other direction, suppose that $S\mu_n(z) \to S\mu(z)$ for all $z \in H_+$. We want to show that

$$\left| \int_{\mathbb{R}} \phi(x) d\mu_n(x) - \int_{\mathbb{R}} \phi(x) d\mu(x) \right| \to 0.$$

Let $\phi \in C_0(\mathbb{R})$ and $\varepsilon > 0$ be fixed, and let M and K be constants such that $|\phi(x)| \le M$ for any x and $|\phi(x)| \le \varepsilon$ for $|x| > K$. Let $f_y(t) = y/\pi(t^2+y^2)$ as in the proof of the previous lemma. Then for small enough $y > 0$, we have $|\phi * f_y(x) - \phi(x)| < \varepsilon$ for all x by the considerations in 1.4. This means that for any n, we have

$$\left| \int_{\mathbb{R}} \phi(x)\, d\mu_n(x) - \int_{\mathbb{R}} \phi * f_y(x)\, d\mu_n(x) \right| \le C\varepsilon \tag{2.8}$$

and the same inequality holds if we replace μ_n with μ.

As in the previous lemma, denote the imaginary parts of functions $x \to \frac{1}{\pi}S\mu(x+iy)$ and $x \to \frac{1}{\pi}S\mu_n(x+iy)$ by g_y and $g_{y,n}$, respectively. We can use Fubini's theorem and the fact that f_y is odd to see that $\int_{\mathbb{R}} \phi * f_y(x)\,d\mu(x) = \int \phi(x)g_y(x)\,dx$ and $\int_{\mathbb{R}} \phi * f_y(x)\,d\mu_n(x) = \int \phi(x)g_{y,n}(x)\,dx$.

By hypothesis $g_{y,n}(x)$ converges to $g_y(x)$ for all $x \in \mathbb{R}$. All values $\phi(x)g_{y,n}(x)$ are dominated by $\frac{M}{y}\chi_{[-K,K]}(x)$ for $|x| \le K$, so by Lebesgue's dominated convergence theorem, we can choose N such that

$$\left|\int_{-K}^{K} \phi(x)g_{y,n}(x)\,dx - \int_{-K}^{K} \phi(x)g_y(x)\,dx\right| \le \varepsilon \tag{2.9}$$

for any $n > N$. For $|x| > K$ we can bound the difference of the integrals by

$$\int_{|x|>K} |\phi(x)|\,|g_{y,n}(x) - g_y(x)|\,dx \le \varepsilon \int_{\mathbb{R}} |g_{y,n}(x) - g_y(x)|\,dx \le \varepsilon R \tag{2.10}$$

for some constant R, since we can see by Fubini's theorem that

$$\int_{\mathbb{R}} |g_{y,n}(x)|\,dx \le \int_{\mathbb{R}}\int_{\mathbb{R}} |f_y(x-t)|\,d\mu(t)\,dx = \int_{\mathbb{R}}\int_{\mathbb{R}} |f_y(x-t)|\,dx\,d\mu(t).$$

Now we can combine the estimates 2.8, 2.9 and 2.10 with the triangle inequality to see that for any $n > N$

$$\begin{aligned}\left|\int_{\mathbb{R}} \phi(x)\,d\mu_n(x) - \int_{\mathbb{R}} \phi(x)\,d\mu(x)\right| \le\ & \left|\int_{\mathbb{R}} \phi(x)\,d\mu_n(x) - \int_{\mathbb{R}} \phi * f_y(x)\,d\mu_n(x)\right| \\ & + \left|\int_{-K}^{K} \phi(x)g_{y,n}(x)\,dx - \int_{-K}^{K} \phi(x)g_y(x)\,dx\right| \\ & + \left|\int_{|x|>K} \phi(x)g_{y,n}(x)\,dx - \int_{|x|>K} \phi(x)g_y(x)\,dx\right| \\ & + \left|\int_{\mathbb{R}} \phi * f_y(x)\,d\mu(x) - \int_{\mathbb{R}} \phi(x)\,d\mu(x)\right| \\ \le\ & \varepsilon(2C + 1 + R).\end{aligned}$$

As ε was arbitrary, we have that $\int_{\mathbb{R}} \phi(x)d\mu_n(x)$ converges to $\int_{\mathbb{R}} \phi(x)d\mu(x)$ almost surely. □

We also need a probabilistic version of the previous theorem to be able to speak about the convergence of random measures.

Theorem 2.11. (Stieltjes continuity theorem, probabilistic version.) *Let $(\mu_n)_{n\in Z_+}$ be a sequence of random finite Borel measures that are surely uniformly bounded by $C < \infty$ and μ a deterministic finite Borel measure on the real line. Then following three are equivalent:*

(1) μ_n converges to μ weakly almost surely.

(2) For any $z \in H_+$ the Stieltjes transform $S\mu_n(z)$ converges to $S\mu(z)$ almost surely.

(3) Almost surely, for all $z \in H_+$ the Stieltjes transform $S\mu_n(z)$ converges to $S\mu(z)$

Proof. The proof of the previous theorem already gives (1) $\implies$ (2) and (3) $\implies$ (1). It remains to show (2) $\implies$ (3). It is enough to show that almost surely, the Stieltjes transform $S\mu_n(z)$ converges to $S\mu(z)$ for all z in the rectangles $R_K = [-K, K] \times [1/K, K]$, as the upper half-plane is a countable union of such rectangles.

Fix K and let (q_i) be an enumeration of the rational points of R_K. Then for all i $S\mu_n(q_i)$ converges to $S\mu(q_i)$ almost surely. Consider the case that every $S\mu_n(q_i)$ converges to $S\mu(q_i)$. Because there are only countably many points q_i, this happens almost surely. Fix $z \in R_K$ and $\varepsilon > 0$. We are going to show that $S\mu_n(z)$ converges to $S\mu(z)$.

Recall that the derivative of any Stieltjes transform in R_K is bounded by CK^2, and therefore $S\mu_n$, $S\mu$ are CK^2-Lipschitz. As the rational points are dense in R_K, we can find a q_i with $|z - q_i| < \varepsilon/(3CK^2)$. Since we assume that $S\mu_n(q_i)$ converges to $S\mu(q_i)$, we know that for large enough n $|S\mu_n(q_i) - S\mu(q_i)| < \varepsilon/3$. Combining these estimates gives for large enough n

$$\begin{aligned}
|S\mu_n(z) - S\mu(z)| &= |S\mu_n(z) - S\mu_n(q_i) + S\mu_n(q_i) - S\mu(q_i) + S\mu(q_i) - S\mu(z)| \\
&\leq |S\mu_n(z) - S\mu_n(q_i)| + |S\mu_n(q_i) - S\mu(q_i)| + |S\mu(q_i) - S\mu(z)| \\
&\leq K^2|z - q_i| + \frac{\epsilon}{3} + K^2|q_i - z| \leq \frac{K^2\varepsilon}{3K^2} + \frac{\epsilon}{3} + \frac{K^2\varepsilon}{3K^2} = \varepsilon
\end{aligned}$$

and it follows that $S\mu_n(z)$ converges to $S\mu(z)$. □

Observe that we only needed the convergence of the imaginary part of the Stieltjes transform to control the convergence of the probability measures. This is simply because the imaginary part of the integral kernel is an approximation of identity. [1]

[1] Considering the limit of the real part of the Stieltjes transform leads to the theory of Hilbert transform which is not described here.

Chapter 3

Deriving the semicircular law

The objective is to derive the almost sure convergence of empirical spectral distributions. In the previous chapter, we saw this to be equivalent to the almost sure convergence of their Stieltjes transforms. In addition, we can also derive the limit probability distribution by Theorem 2.6.

If we denote the n eigenvalues of W_n by $\lambda_1, \ldots, \lambda_n$, the Stieltjes transform of the ESD can be written as

$$\begin{aligned} S\mu_{\frac{1}{\sqrt{n}}W_n}(z) &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda_i/\sqrt{n}-z} \\ &= \frac{1}{n}tr(\frac{1}{\sqrt{n}}W_n - zI_n)^{-1}. \end{aligned}$$

For the rest of this chapter, z will be a fixed complex number with positive imaginary part. We will prove that the Stieltjes transforms of the empirical spectral distributions converge almost surely in two parts. First we prove that the expectations of the Stieltjes transforms converge, then we prove that the difference of Stieltjes transform and its expectation goes to zero almost surely. We start by considering the expectation of the empirical spectral measure. By linearity of expectation, we have

$$\begin{aligned} ES\mu_{\frac{1}{\sqrt{n}}W_n}(z) &= E(\frac{1}{n}tr(\frac{1}{\sqrt{n}}W_n - zI_n)^{-1}) = E(\frac{1}{n}\sum_{l=1}^{n}(\frac{1}{\sqrt{n}}W_n - zI_n)^{-1}_{ll}) \\ &= \frac{1}{n}\sum_{l=1}^{n}E(\frac{1}{\sqrt{n}}W_n - zI_n)^{-1}_{ll} = E(\frac{1}{\sqrt{n}}W_n - zI_n)^{-1}_{nn}. \end{aligned}$$

The last equality holds because by Cramer's rule (Theorem C.1), the random variables $(\frac{1}{\sqrt{n}}W_n - zI_n)^{-1}_{ll}$ are identically distributed. Using Theorem C.2, we have

$$E\left(\frac{1}{\sqrt{n}}W_n - zI_n\right)^{-1}_{nn} = E\frac{1}{-z + \frac{\xi_{nn}}{\sqrt{n}} - \frac{1}{n}C_n^*\left(\frac{1}{\sqrt{n}}W_{n-1} - zI_{n-1}\right)^{-1}C_n},$$

where C_n is the column vector $(\xi_{in})_{i=1}^{n-1}$. First, consider the term $\xi_{nn}/\sqrt{n}$. From zero mean and unit variance hypothesis, it immediately follows by Markov's inequality that

$$P\left(\left|\frac{\xi_{nn}}{\sqrt{n}}\right| > \frac{\lambda}{\sqrt[4]{n}}\right) \leq \frac{1}{\sqrt{n}\lambda^2}.$$

This means informally that the term contributes to the expectation only by $o(1)$. Notice that we could have assumed instead that the diagonal elements have a finite mean and variance, and a similar result would have followed.

Now consider the other random term. Notice that the matrix $A_n = (\frac{1}{\sqrt{n}}W_{n-1} - zI_{n-1})^{-1}$ has $n-1$ eigenvalues, and the eigenvalues are bounded below in absolute value by $|\operatorname{Im}(z)|^{-1}$. Also, matrix A_n is independent from the vector C_n. We will prove the concentration of the term $C_n^*A_nC_n$ near a deterministic value.

3.1 Concentration of the quadratic form $C_n^*A_nC_n$

In the previous section the matrix A_n is independent of the vector C_n. We will consider the behaviour of the term $C_n^*A_nC_n$ for any fixed A_n and use independence of C_n and A_n to derive estimates from this.

Lemma 3.1. *For a deterministic $(n-1)\times(n-1)$-matrix B, we have*

$$E(C_n^*BC_n) = tr(B).$$

Proof. Denote $c_i = \xi_{in}$ for $1 \leq i \leq n-1$. As the random variables c_i are independent and have mean zero and unit variance, we may use $E\overline{c_i}c_j = \delta_{ij}$ to compute

$$E(C_n^*BC_n) = E(\sum_{i=1}^{n-1}\sum_{j=1}^{n-1}\overline{c_i}b_{ij}c_j) = \sum_{i=1}^{n-1}(b_{ii}E(\overline{c_i}c_i) + \sum_{j=1,j\neq i}^{n-1}b_{ij}E(\overline{c_i}c_j)) = tr(B).$$

□

Lemma 3.2. *For a deterministic $(n-1)\times(n-1)$ matrix B we have the following estimate for the variance:*

$$Var(C_n^*BC_n) \leq R\|B\|_2^2 \leq nR\|B\|^2$$

for some constant $R > 0$ independent from B. Here $\|B\|_2$ denotes the Hilbert-Schmidt norm of B and $\|B\|$ is the operator norm of B. In other words,

$$\|B\|_2^2 := \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} |b_{ij}|^2$$

and

$$\|B\| := \sup_{x \in \mathbb{C}^{n-1} : |x|=1} |Bx|.$$

Recall that the variance of the random variable X is defined as

$$Var(X) = E|X|^2 - |EX|^2.$$

Proof. We computed the expectation in the previous lemma. As $Var(Y) = E(|Y|^2) - |E(Y)|^2$, we need to compute $E(|C_n^* B C_n|^2)$. Similarly to the previous lemma, we denote the components of C_n by c_i and compute

$$\begin{aligned} E(|C_n^* B C_n|^2) &= E(|\sum_{i=1}^{n-1} \sum_{j=1}^{n-1} \overline{c_i} b_{ij} c_j|^2) = E((\sum_{i=1}^{n-1} \sum_{j=1}^{n-1} \overline{c_i} b_{ij} c_j)\overline{(\sum_{i=1}^{n-1} \sum_{j=1}^{n-1} \overline{c_i} b_{ij} c_j)}) \\ &= \sum_{i_1=1}^{n-1} \sum_{i_2=1}^{n-1} \sum_{j_1=1}^{n-1} \sum_{j_2=1}^{n-1} E(\overline{c_{i_1}} b_{i_1 j_1} c_{j_1} c_{i_2} \overline{b_{i_2 j_2} c_{j_2}}). \end{aligned}$$

Now recall that all c_k have mean zero and are independent. This means that if in one term one of the indices i_1, i_2, j_1 and j_2 is distinct from all the others, that term vanishes. The only remaining terms are the ones in one of the following cases:

(1) $i_1 = i_2 = j_1 = j_2$
(2) $i_1 = i_2 \neq j_1 = j_2$
(3) $i_1 = j_1 \neq i_2 = j_2$
(4) $i_1 = j_2 \neq i_2 = j_1$

Let's go through these cases individually. Denote by S_i the sum of the terms from case i. For case (1), recall that the non-diagonal elements have a finite fourth moment, so $E(|c_k|^4) = K < \infty$. Letting $i_1 = i_2 = j_1 = j_2 = k$, we have

$$S_1 := \sum_{k=1}^{n-1} E(\overline{c_k} b_{kk} c_k c_k \overline{b_{kk} c_k}) = \sum_{k=1}^{n-1} |b_{kk}|^2 E(|c_k|^4) = \sum_{k=1}^{n-1} K |b_{kk}|^2$$

In case (2), set $i_1 = i_2 = k_1$ and $j_1 = j_2 = k_2$, and we have

$$\begin{aligned} S_2 &:= \sum_{k_1=1}^{n-1} \sum_{k_2=1, k_1 \neq k_2}^{n-1} E(\overline{c_{k_1}} b_{k_1 k_2} c_{k_2} c_{k_1} \overline{b_{k_1 k_2} c_{k_2}}) = \sum_{k_1=1}^{n-1} \sum_{k_2=1, k_1 \neq k_2}^{n-1} |b_{k_1 k_2}|^2 E(|c_{k_1}|^2) E(|c_{k_2}|^2) \\ &= \sum_{k_1=1}^{n-1} \sum_{k_2=1, k_1 \neq k_2}^{n-1} |b_{k_1 k_2}|^2 \end{aligned}$$

In case (3), setting $i_1 = j_1 = k_1$ and $i_2 = j_2 = k_2$ gives

$$\begin{aligned} S_3 &:= \sum_{k_1=1}^{n-1} \sum_{k_2=1, k_1 \neq k_2}^{n-1} E(\overline{c_{k_1}} b_{k_1 k_1} c_{k_1} c_{k_2} \overline{b_{k_2 k_2} c_{k_2}}) \\ &= \sum_{k_1=1}^{n-1} \sum_{k_2=1, k_1 \neq k_2}^{n-1} b_{k_1 k_1} \overline{b_{k_2 k_2}} E(|c_{k_1}|^2) E(|c_{k_2}|^2) = \sum_{k_1=1}^{n-1} \sum_{k_2=1, k_1 \neq k_2}^{n-1} b_{k_1 k_1} \overline{b_{k_2 k_2}} \end{aligned}$$

Notice that in the sums (1) and (2) all the terms are real, and case (3) may be written as

$$S_3 = \sum_{k_1=1}^{n-1} \sum_{k_2=k_1+1}^{n-1} (b_{k_1 k_1} \overline{b_{k_2 k_2}} + \overline{b_{k_1 k_1}} b_{k_2 k_2}) \in \mathbb{R}.$$

Finally in case (4), setting $i_1 = j_2 = k_1$ and $i_2 = j_1 = k_2$ to compute

$$S_4 = \sum_{k_1=1}^{n-1} \sum_{k_2=1, k_1 \neq k_2}^{n-1} E(\overline{c_{k_1}} b_{k_1 k_2} c_{k_2} c_{k_2} \overline{b_{k_2 k_1} c_{k_1}}) = \sum_{k_1=1}^{n-1} \sum_{k_2=1, k_1 \neq k_2}^{n-1} b_{k_1 k_2} \overline{b_{k_2 k_1}} E(\overline{c_{k_1}}^2) E(c_{k_2}^2)$$

We know that S_4 must be real, which also follows by symmetry considerations. Using the triangle inequality and the arithmetic-geometric mean inequality, we estimate

$$\begin{aligned} S_4 &= \sum_{k_1=1}^{n-1} \sum_{k_2=1, k_1 \neq k_2}^{n-1} b_{k_1 k_2} \overline{b_{k_2 k_1}} E(\overline{c_{k_1}}^2) E(c_{k_2}^2) \\ &\leq \sum_{k_1=1}^{n-1} \sum_{k_2=1, k_1 \neq k_2}^{n-1} |b_{k_1 k_2}| |b_{k_2 k_1}| |E(\overline{c_{k_1}}^2)| |E(c_{k_2}^2)| \\ &\leq \sum_{k_1=1}^{n-1} \sum_{k_2=1, k_1 \neq k_2}^{n-1} |b_{k_1 k_2}| |b_{k_2 k_1}| E(|c_{k_1}|^2) E(|c_{k_2}|^2) = \sum_{k_1=1}^{n-1} \sum_{k_2=1, k_1 \neq k_2}^{n-1} |a_{k_1 k_2}| |b_{k_2 k_1}| \\ &\leq \sum_{k_1=1}^{n-1} \sum_{k_2=1, k_1 \ neq k_2}^{n-1} \frac{|b_{k_1 k_2}|^2 + |b_{k_2 k_1}|^2}{2} = \sum_{k_1=1}^{n-1} \sum_{k_2=1, k_1 \neq k_2}^{n-1} |b_{k_1 k_2}|^2. \end{aligned}$$

Now

$$\begin{aligned}
Var(C_n^* B C_n) &= S_1 + S_2 + S_3 + S_4 - |E(C_n^* B C_n)|^2 \\
&= \sum_{k=1}^{n-1} (K-1)|b_{kk}|^2 + |tr(B)|^2 + S_2 + S_4 - |tr(B)|^2 \\
&= \sum_{k=1}^{n-1} (K-1)|b_{kk}|^2 + \sum_{k_1=1}^{n-1} \sum_{k_2=1, k_1 \neq k_2}^{n-1} |b_{k_1 k_2}|^2 + S_4 \\
&\leq \sum_{k=1}^{n-1} (K-1)|b_{kk}|^2 + \sum_{k_1=1}^{n-1} \sum_{k_2=1, k_1 \neq k_2}^{n-1} 2|b_{k_1 k_2}|^2 \\
&\leq \max(K-1, 2) \sum_{k_1=1}^{n-1} \sum_{k_2=1}^{n-1} |b_{k_1 k_2}|^2 = R\|B\|_2^2.
\end{aligned}$$

Observe that the constant $R = \max(K-1, 2)$ does not depend on n or B. We know that then $|Be_i|^2 \leq \|B\|^2$. But this means that

$$R\|B\|_2^2 \;=\; R \sum_{k_1=1}^{n-1} \sum_{k_2=1}^{n-1} |b_{k_1 k_2}|^2 = R \sum_{k_2=1}^{n-1} \sum_{k_1=1}^{n-1} |b_{k_1 k_2}|^2 = R \sum_{k_2=1}^{n-1} |Be_{k_2}|^2 \leq Rn\|B\|^2$$

$\square$

We can apply Chebyshev's inequality to conclude that

$$P(|C_n^* B C_n - tr(B)| > \lambda) \leq nR\|B\|^2/\lambda^2,$$

or equivalently

$$P(|C_n^* B C_n - tr(B)| > n^{\frac{3}{4}}\lambda) \leq R\|B\|^2/\sqrt{n}\lambda^2.$$

This implies that if we condition the matrix A_n to be fixed, we can estimate the conditional probability as

$$P(|\frac{1}{n} C_n^* A_n C_n - \frac{1}{n} tr(A_n)| > \lambda/\sqrt[4]{n} | A_n) \leq R\|A_n\|^2/\sqrt{n}\lambda^2.$$

The matrix $A_n = (\frac{1}{\sqrt{n}} W_{n-1} - zI_{n-1})^{-1}$ is an inverse of an $(n-1)\times(n-1)$-matrix with $n-1$ eigenvalues that have absolute value over y by spectral theorem. This means that $\|A_n\| \leq 1/y$ for all possible A_n. As A_n and C_n are independent, we see that unconditionally

$$P\left(\left| \frac{1}{n} C_n^* A_n C_n - \frac{1}{n} tr(A_n) \right| > \lambda/\sqrt[4]{n} \right) \leq R/y^2\sqrt{n}\lambda^2. \tag{3.3}$$

Next, we will approximate the term $\frac{1}{n} tr(A_n)$ by the original Stieltjes transform $S\mu_{\frac{1}{\sqrt{n}} W_n}(z)$.

Lemma 3.4. *For any l, denote by W_n^l the matrix W_n with lth row and column removed. We have surely the estimate*

$$|\frac{1}{n}tr((\frac{1}{\sqrt{n}}W_n^l - zI_{n-1})^{-1}) - S_{\mu_{\frac{1}{\sqrt{n}}W_n}}(z)| \leq \frac{C_z}{n}.$$

Proof. Knowing that both W_n and W_n^l are hermitian matrices, let $\lambda_1 \leq \lambda_2 \leq \dots \lambda_n$ be the eigenvalues of W_n and $\lambda_1' \leq \lambda_2' \leq \dots \lambda_{n-1}'$ be the eigenvalues of W_n^l. We know by Cauchy interlacing formula (Theorem B.3) that

$$\lambda_i \leq \lambda_i' \leq \lambda_{i+1}(A)$$

for $1 \leq i \leq n-1$. We see that the eigenvalues of the matrix $B = (\frac{1}{\sqrt{n}}W_n^l - zI_{n-1})^{-1}$ are $(\lambda_i'/\sqrt{n} - z)^{-1}$ for $1 \leq i \leq n-1$. Rewriting the difference we want to estimate and using triangle inequality gives

$$\begin{aligned}
\left|\frac{1}{n}tr(B) - S_{\mu_{\frac{1}{\sqrt{n}}W_n}}(z)\right| &= \frac{1}{n}\left|\sum_{i=1}^{n-1}\frac{1}{\frac{\lambda_i'}{\sqrt{n}} - z} - \sum_{i=1}^{n}\frac{1}{\frac{\lambda_i}{\sqrt{n}} - z}\right| \\
&\leq \frac{1}{n}\left(\sum_{i=1}^{n-1}\left|\frac{1}{\frac{\lambda_i'}{\sqrt{n}} - z} - \frac{1}{\frac{\lambda_i}{\sqrt{n}} - z}\right| + \left|\frac{1}{\frac{\lambda_n}{\sqrt{n}} - z}\right|\right).
\end{aligned}$$

We note that the intervals $]\lambda_i/\sqrt{n}, \lambda_i'/\sqrt{n}[$ are disjoint by Cauchy interlacing formula. For any sequence $t_1 \leq t_2 \leq \dots \leq t_{2m-1} \leq t_{2m}$, we can estimate with triangle inequality

$$\begin{aligned}
\sum_{j=1}^{m}\left|\frac{1}{t_{2j} - z} - \frac{1}{t_{2j-1} - z}\right| &= \sum_{j=1}^{m}\left|\frac{1}{t_{2j} - (x+iy)} - \frac{1}{t_{2j-1} - (x+iy)}\right| \\
&= \sum_{j=1}^{m}\left|\frac{t_{2j} - x + iy}{t_{2j} - (x+iy)} - \frac{t_{2j-1} - x + iy}{(t_{2j-1} - x)^2 + y^2}\right| \\
&\leq \sum_{j=1}^{m}\left|\frac{t_{2j} - x}{(t_{2j} - x)^2 + y^2} - \frac{t_{2j-1} - x}{(t_{2j-1} - x)^2 + y^2}\right| \\
&\quad + \sum_{j=1}^{m}\left|\frac{y}{(t_{2j} - x)^2 + y^2} - \frac{y}{(t_{2j-1} - x)^2 + y^2}\right|.
\end{aligned}$$

Now we see that function $x \to (x-a)/((x-a)^2 + b^2)$ is continuously differentiable. It has a maximum $1/2b$ at the point $x = a+b$, a minimum $-1/2b$ at $x = a-b$ and it approaches zero at $\pm\infty$. This means that the first sum is at most $2/b$. Likewise, considering the derivative of the function $x \to b/((x-a)^2+b^2)$ we see that it has a maximum $1/b$ at the point $x = a$ and it approaches zero at $\pm\infty$, so the second sum is at most $2/b$.

We thus see that

$$\begin{aligned}\left|\frac{1}{n}tr(\frac{1}{\sqrt{n}}W_n^l - zI_{n-1})^{-1} - S_{\mu_{\frac{1}{\sqrt{n}}W_n}}(z)\right| &\leq \frac{1}{n}\sum_{i=1}^{n-1}\left|\frac{1}{\frac{\lambda_i'}{\sqrt{n}}-z} - \frac{1}{\frac{\lambda_i}{\sqrt{n}}-z}\right| + \left|\frac{1}{\frac{\lambda_n}{\sqrt{n}}-z}\right| \\ &\leq \frac{1}{n}\left(\frac{2}{b}+\frac{2}{b}+\left|\frac{1}{\frac{\lambda_n}{\sqrt{n}}-z}\right|\right) \leq \frac{1}{n}\left(\frac{4}{b}+\frac{1}{b}\right) = \frac{5}{nb}.\end{aligned}$$

□

Finally, the following special case of *McDiarmid's inequality*[7] allows us to estimate the Stieltjes transform by its expectation:

Lemma 3.5. *For any index n and positive number λ, we have*

$$P(|S\mu_{\frac{1}{\sqrt{n}}W_n}(z) - ES\mu_{\frac{1}{\sqrt{n}}W_n}(z)| > \lambda) \leq Ce^{-cn\lambda^2}.$$

for some constants $C, c > 0$, depending only on z.

Proof. We define the real random variables D_k for $0 \leq k \leq n$ to be the conditional expectation

$$D_k = \mathrm{Re}(E(S\mu_{\frac{1}{\sqrt{n}}W_n}(z)|(\xi_{ij})_{i\leq j\leq k}),$$

in other words, the conditional expectation of the real part of the Stieltjes transform when we fix the random variables ξ_{ij} for i, j at most k. [1] For these D_k, we observe that $D_0 = \mathrm{Re}(ES\mu_{\frac{1}{\sqrt{n}}W_n}(z))$ and $D_n = \mathrm{Re}(S\mu_{\frac{1}{\sqrt{n}}W_n}(z))$. We claim that the inequality $|D_k - D_{k-1}| \leq C_z/n$ holds surely. To see this, note that

$$D_k - D_{k-1} = \mathrm{Re}\,E(S\mu_{\frac{1}{\sqrt{n}}W_n}(z) - E(S\mu_{\frac{1}{\sqrt{n}}W_n}(z)|(\xi_{ij})_{i\leq j\neq k})|(\xi_{ij})_{i\leq j\leq k}).$$

But we had in Lemma 3.4 that surely

$$|\frac{1}{n}tr((\frac{1}{\sqrt{n}}W_n^k - zI_{n-1})^{-1}) - S_{\mu_{\frac{1}{\sqrt{n}}W_n}}(z)| \leq \frac{C_z}{n},$$

[1]This construction is known as the Doob martingale [8]. The argument we will do is essentially using Azuma's inequality [9] for this sequence.

so $|S\mu_{\frac{1}{\sqrt{n}}W_n}(z) - E(S\mu_{\frac{1}{\sqrt{n}}W_n}(z)|(\xi_{ik})_{i\le k})|$ is bounded above by C_z/n. Taking the conditional expectation gives $|D_k - D_{k-1}| \le C_z/n$. Next, we consider the exponential moment Ee^{tD_n}. We are going to prove next that

$$Ee^{tD_k} \le e^{\frac{1}{8}t^2\left(\frac{C_z}{n}\right)^2} Ee^{tD_k - 1} \tag{3.6}$$

for any $1 \le k \le n$. Applied inductively, this gives

$$Ee^{tD_n} \le e^{\frac{C_z^2 t^2}{8n}} Ee^{tD_0}.$$

To start the proof of (3.6), we note that D_{k-1} does not depend on ξ_{ik} for $i \le k$. This means that

$$\begin{aligned} Ee^{tD_k} &= EE(e^{tD_k}|(\xi_{ij})_{i\le j\neq k}) \\ &= EE(e^{t(D_k-D_{k-1})}e^{tD_{k-1}}|(\xi_{ij})_{i\le j\neq k}) \\ &= E(e^{tD_{k-1}}E(e^{t(D_k-D_{k-1})}|(\xi_{ij})_{i\le j\neq k})). \end{aligned}$$

Using Hoeffding's lemma (Theorem A.2), we conclude

$$E(e^{t(D_k-D_{k-1})}|(\xi_{ij})_{i\le j\neq k}) \le e^{\frac{1}{8}t^2\left(\frac{C_z}{n}\right)^2},$$

so we have

$$Ee^{tD_k} = E(e^{tD_{k-1}}E(e^{t(D_k-D_{k-1})}|(\xi_{ij})_{i\le j\neq k})) \le e^{\frac{1}{8}t^2\left(\frac{C_z}{n}\right)^2} Ee^{tD_{k-1}},$$

i.e. (3.6). We have

$$Ee^{t\,\mathrm{Re}(S\mu_{\frac{1}{\sqrt{n}}W_n}(z))} \le e^{\frac{C_z^2 t^2}{8n}} e^{tE\,\mathrm{Re}(S\mu_{\frac{1}{\sqrt{n}}W_n}(z))},$$

or equivalently

$$Ee^{t(\mathrm{Re}(S\mu_{\frac{1}{\sqrt{n}}W_n}(z)) - E\,\mathrm{Re}(S\mu_{\frac{1}{\sqrt{n}}W_n}(z)))} \le e^{\frac{C_z^2 t^2}{8n}}.$$

For any $a > 0$, we have by Markov's inequality

$$P\left(e^{t(\mathrm{Re}(S\mu_{\frac{1}{\sqrt{n}}W_n}(z)) - E\,\mathrm{Re}(S\mu_{\frac{1}{\sqrt{n}}W_n}(z)))} \ge a\right) \le \frac{e^{\frac{C_z^2 t^2}{8n}}}{a},$$

which we may rewrite setting $a = e^{t\lambda}$ as

$$P\left((\mathrm{Re}(S\mu_{\frac{1}{\sqrt{n}}W_n}(z)) - E\,\mathrm{Re}(S\mu_{\frac{1}{\sqrt{n}}W_n}(z))) \ge \lambda\right) \le \frac{e^{\frac{C_z^2 t^2}{8n}}}{e^{t\lambda}}.$$

The right-hand side has a minimum at $t = (4\lambda n)/(C_z^2)$, so we get

$$P\left((\mathrm{Re}(S\mu_{\frac{1}{\sqrt{n}}W_n}(z)) - E\,\mathrm{Re}(S\mu_{\frac{1}{\sqrt{n}}W_n}(z))) \geq \lambda\right) \leq e^{\frac{-2n\lambda^2}{C_z^2}}.$$

If we do the same argument with random variables $-D_k$, we obtain

$$P\left((\mathrm{Re}(S\mu_{\frac{1}{\sqrt{n}}W_n}(z)) - E\,\mathrm{Re}(S\mu_{\frac{1}{\sqrt{n}}W_n}(z))) \leq -\lambda\right) \leq e^{\frac{-2n\lambda^2}{C_z^2}}.$$

Taken together, these two estimates imply that

$$P\left(|\,\mathrm{Re}(S\mu_{\frac{1}{\sqrt{n}}W_n}(z) - ES\mu_{\frac{1}{\sqrt{n}}W_n}(z))| \geq \lambda\right) \leq 2e^{\frac{-2n\lambda^2}{C_z^2}}.$$

We can repeat this argument for the imaginary part to conclude

$$P\left(|\,\mathrm{Im}(S\mu_{\frac{1}{\sqrt{n}}W_n}(z) - ES\mu_{\frac{1}{\sqrt{n}}W_n}(z))| \geq \lambda\right) \leq 2e^{\frac{-2n\lambda^2}{C_z^2}}.$$

This means that we have an upper bound

$$\begin{aligned}
P(|S\mu_{\frac{1}{\sqrt{n}}W_n}(z) - ES\mu_{\frac{1}{\sqrt{n}}W_n}(z)| > \lambda) &\leq P(|\,\mathrm{Re}(S\mu_{\frac{1}{\sqrt{n}}W_n}(z) - ES\mu_{\frac{1}{\sqrt{n}}W_n}(z))| > \lambda/2) \\
&\quad + P(|\,\mathrm{Im}(S\mu_{\frac{1}{\sqrt{n}}W_n}(z) - ES\mu_{\frac{1}{\sqrt{n}}W_n}(z))| > \lambda/2) \\
&\leq 2e^{\frac{-n\lambda^2}{2C_z^2}} + 2e^{\frac{-n\lambda^2}{2C_z^2}} \\
&= 4e^{\frac{-n\lambda^2}{2C_z^2}}.
\end{aligned}$$

□

We can combine estimate (3.3) with the estimates from Lemma 3.4 and Lemma 3.5 to conclude by the triangle inequality

$$\begin{aligned}
P(|\frac{1}{n}C_n^* A_n C_n - ES\mu_{\frac{1}{\sqrt{n}}W_n}(z)| \geq 3/\sqrt[4]{n}) &\leq P(|\frac{1}{n}C_n^* A_n C_n - \frac{1}{n}tr(A_n)| \geq 1/\sqrt[4]{n}) \\
&\quad + P(|\frac{1}{n}tr(A_n) - S\mu_{\frac{1}{\sqrt{n}}W_n}(z)| \geq 1/\sqrt[4]{n}) \\
&\quad + P(|S\mu_{\frac{1}{\sqrt{n}}W_n}(z) - ES\mu_{\frac{1}{\sqrt{n}}W_n}(z)| \geq 1/\sqrt[4]{n}) \\
&\leq \frac{R}{y^2\sqrt{n}} + 0 + Ce^{-c\sqrt{n}}
\end{aligned}$$

for any n such that $1/\sqrt[4]{n} > C_z/n$, meaning $n > C_z^{\frac{3}{4}}$.

Remark 3.7. If we look at the proofs of the concentration results used in previous section, we can actually generalize them easily so that the assumption on elements being identically distributed is not necessary. While the elements of the diagonal of the inverse matrix are no longer identically distributed, we can permute the rows and columns to obtain uniform bounds for expectations of the diagonal elements, receiving the same result. We do not need this result for the rest of the work.

3.2 Finishing the computation

We had the equation

$$ES\mu_{\frac{1}{\sqrt{n}}W_n}(z) = E\frac{-1}{z - \frac{\xi_{nn}}{\sqrt{n}} + \frac{1}{n}C_n^*\left(\frac{1}{\sqrt{n}}W_{n-1} - zI_{n-1}\right)^{-1}C_n}.$$

The computations thus far have shown that for the denominator we have

$$\left| z - \frac{\xi_{nn}}{\sqrt{n}} + \frac{1}{n}C_n^*\left(\frac{1}{\sqrt{n}}W_{n-1} - zI_{n-1}\right)^{-1}C_n - \left(z + ES\mu_{\frac{1}{\sqrt{n}}W_n}(z)\right)\right| < o(1)$$

with probability $1 - o(1)$. As the denominator is bounded away from zero, we have by continuity

$$ES\mu_{\frac{1}{\sqrt{n}}W_n}(z) = -\frac{1}{z + ES\mu_{\frac{1}{\sqrt{n}}W_n}(z)} + o(1).$$

But this is actually a second order polynomial equation for $ES\mu_{\frac{1}{\sqrt{n}}W_n}(z)$ for a fixed z. We can solve this and use the fact that the branches of the complex square root are continuous away from zero to see that

$$ES\mu_{\frac{1}{\sqrt{n}}W_n}(z) = \frac{-z \pm \sqrt{z^2-4}}{2} + o(1).$$

Here the branch of the square root could depend on n. But since $S\mu$ maps the upper half-plane to the upper half-plane,

$$ES\mu_{\frac{1}{\sqrt{n}}W_n}(z) = \frac{-z + \sqrt{z^2-4}}{2} + o(1),$$

where the branch of the square root has positive imaginary part in upper half-plane.

We have shown that the expectations of the Stieltjes transforms converge to a limit function. Recall that we proved the following estimate as Lemma 3.4.

$$P(|S\mu_{\frac{1}{\sqrt{n}}W_n}(z) - ES\mu_{\frac{1}{\sqrt{n}}W_n}(z)| > \lambda) \le Ce^{-cn\lambda^2}.$$

For any $\varepsilon > 0$, the sum $\sum_{n=1}^{\infty} Ce^{-cn\varepsilon}$ converges as a geometric series. Combining this estimate with Borel-Cantelli lemma (Theorem A.1) implies that almost surely

$$S\mu_{\frac{1}{\sqrt{n}}W_n}(z) \to g(z) = \frac{-z + \sqrt{z^2 - 4}}{2}.$$

We still need to find a probability measure μ with $S\mu = g$. Let us investigate the behaviour of the imaginary part of g near the real axis.

$$\begin{aligned} \frac{1}{\pi}\operatorname{Im}(g(x+yi)) &= \frac{y}{2\pi} + \frac{\sqrt{\frac{4-x^2+y^2+\sqrt{(x^2-y^2-4)^2+4x^2y^2}}{2}}}{2\pi} \\ &\to \frac{1}{2\pi}\sqrt{\max(0, 4-x^2)}, \end{aligned}$$

i.e. the limit of the imaginary part of the function g is the *Wigner semicircular distribution* μ_{sc}. Now that we know what the limiting probability measure should be, the problem of verifying the limit measure reduces to straightforward computation.

Theorem 3.8. *The Stieltjes transform of μ_{sc} is the function g we obtained, so*

$$S\mu_{sc}(z) = g(z)$$

for every $z \in H_+$.

Proof. We will compute the Stieltjes transform at point z in the upper half-plane. First, making the substitution $x = 2\cos\theta$, we get

$$\begin{aligned} S\mu_{sc}(z) &= \int_{\mathbb{R}} \frac{1}{x-z} d\mu_{sc}(x) = \frac{1}{2\pi}\int_{-2}^{2} \frac{\sqrt{4-x^2}}{x-z} dx = \frac{1}{2\pi}\int_0^{\pi} \frac{\sqrt{4-4\cos^2\theta}}{2\cos\theta - z} 2\sin\theta d\theta \\ &= \frac{1}{2\pi}\int_0^{\pi} \frac{4\sin^2\theta}{2\cos\theta - z} d\theta = \frac{1}{4\pi}\int_0^{2\pi} \frac{4\sin^2\theta}{2\cos\theta - z} d\theta. \end{aligned}$$

Expressing the sine and cosine via complex exponential function and making substitution $e^{i\theta} = w$, we get

$$\begin{aligned} S\mu_{sc}(z) &= \frac{1}{4\pi}\int_0^{2\pi} \frac{4\sin^2\theta}{2\cos\theta - z} d\theta = \frac{1}{4\pi}\int_0^{2\pi} \frac{4(\frac{1}{2i}(e^{i\theta} - e^{-i\theta}))^2}{2(\frac{1}{2}(e^{i\theta} + e^{-i\theta})) - z} d\theta \\ &= \frac{1}{4\pi}\int_0^{2\pi} \frac{-(e^{2i\theta} - 2 + e^{-2i\theta})}{(e^{i\theta} + e^{-i\theta}) - z} d\theta = \frac{1}{4\pi}\int_{|w|=1} \frac{2 - w^2 - w^{-2}}{(w + w^{-1}) - z} \frac{dw}{iw} \\ &= \frac{1}{4\pi i}\int_{|w|=1} \frac{2 - w^2 - w^{-2}}{w^2 + 1 - zw} dw = \frac{1}{4\pi i}\int_{|w|=1} \frac{-w^4 + 2w^2 - 1}{w^2(w^2 + 1 - zw)} dw \\ &= \frac{1}{4\pi i}\int_{|w|=1} h_z(w) dw. \end{aligned}$$

We are going to apply the residue theorem to compute the integral. As the function h_z has poles at $w = 0$ and $w = \frac{z \pm \sqrt{z^2-4}}{2}$, we need to figure out if the poles at $w = \frac{z \pm \sqrt{z^2-4}}{2}$ lie in the unit disk. Denote those two poles by p_1 and p_2. Then, $p_1 + p_2 = z$ and $p_1 p_2 = 1$, which implies that exactly one of them lies in the unit disk. Since z has positive imaginary part, the one with smaller absolute value must have negative imaginary part. Therefore, the pole $w = \frac{z-\sqrt{z^2-4}}{2} = p_1$ lies in the unit disk, where the square root branch has positive imaginary part.

Calculating the residues is straightforward. For $w = 0$:

$$\begin{aligned}
Res(h_z, 0) &= \left(\frac{d}{dw} \frac{-w^4 + 2w^2 - 1}{w^2 - zw + 1} \right)_{w=0} \\
&= \left(\frac{(-4w^3 + 4w)(w^2 - zw + 1) - (-w^4 + 2w^2 - 1)(2w - z)}{(w^2 - zw + 1)^2} \right)_{w=0} \\
&= \frac{(-4 \cdot 0^3 + 4 \cdot 0)(0^2 - z \cdot 0 + 1) - (-0^4 + 2 \cdot 0^2 - 1)(2 \cdot 0 - z)}{(0^2 - z \cdot 0 + 1)^2} \\
&= -z
\end{aligned}$$

For $w = p_1$, we use the fact that $p_1 p_2 = 1$:

$$\begin{aligned}
Res(h_z, p_1) &= \left(\frac{-w^4 + 2w^2 - 1}{w^2(w - p_2)} \right)_{w=p_1} = \frac{-p_1^4 + 2p_1^2 - 1}{p_1^2(p_1 - p_2)} \\
&= \frac{-(p_1^2 - 1)^2}{p_1^2(p_1 - p_2)} = \frac{-(p_1^2 - p_1 p_2)^2}{p_1^2(p_1 - p_2)} = \frac{-p_1^2(p_1 - p_2)^2}{p_1^2(p_1 - p_2)} \\
&= p_2 - p_1 = \frac{z + \sqrt{z^2 - 4}}{2} - \frac{z - \sqrt{z^2 - 4}}{2} = \sqrt{z^2 - 4}.
\end{aligned}$$

By residue theorem we have

$$\begin{aligned}
S\mu_{sc}(z) &= \frac{1}{4\pi i} \int_{|w|=1} \frac{-w^4 + 2w^2 - 1}{w^2(w^2 + 1 - zw)} dw \\
&= \frac{1}{2}(-z + \sqrt{z^2 - 4}).
\end{aligned}$$

□

By Stieltjes continuity theorem, we can deduce that $\mu_{\frac{1}{\sqrt{n}} W_n}$ converges to the semicircular distribution almost surely. We can finally prove our main result.

Theorem 3.9. (Wigner semicircular law.) *For an ensemble of Wigner matrices $(W_n)_{n=1}^{\infty}$ with all elements of matrices being independent and identically distributed, having zero mean, unit variance and bounded fourth moment, the empirical spectral distributions $\mu_{\frac{1}{\sqrt{n}} W_n}$ converges almost surely to Wigner semicircular distribution μ_{sc}.*

Proof. We have proved that for any fixed $z \in H_+$ $S\mu_{\frac{1}{\sqrt{n}}W_n}(z)$ converges to $S\mu_{sc}(z)$ almost surely. This is equivalent to the claim by Stieltjes continuity theorem. $\square$

Let us finish this section with the following remark to illustrate a fundamental difference between the semicircular law and the central limit theorem. Recall that the central limit theorem states that for a sequence of independent identically distributed real random variables (X_i) with zero mean and unit variance, the distribution of $(\sum_{i=1}^{n} X_i)/\sqrt{n}$ approaches the standard normal distribution. As the sums of independent Gaussian random variables are Gaussian, this can be used when proving the central limit theorem. However, the empirical spectral distributions will never have the same kind of stability for any choice of coefficients:

Theorem 3.10. *It is not possible to choose the distribution of the random variables ξ_{ij} in Wigner matrix in a way such that for all n*

$$E\mu_{\frac{1}{\sqrt{n}}W_n} = \mu_{sc}.$$

In fact, for any choice of the random variables, the equality does not hold for any $n > C$ for some constant C independent from the choice. Here $E\mu$ denotes the expectation measure of the random measure μ, i.e. $E\mu(A) = E(\mu(A))$.

This theorem is an immediate consequence of the next lemma.

Lemma 3.11. *For any $K > 0$, we have for sufficiently large n*

$$P(\mu_{\frac{1}{\sqrt{n}}W_n}(\{x \in \mathbb{R} : |x| > K\} > 0)) > 0.$$

Proof. The idea of the proof is to see that if the coefficients in the matrix happen to behave similarly, the matrix W_n will have operator norm larger than cn for some constant $c > 0$, which is certainly too large. We need to be slightly careful with the diagonal elements since we did not assume them to be identically distributed to the off-diagonal elements.

By unit variance hypothesis, we know that $P(|\xi_{ij}| \geq 1) > 0$ for any $i < j$. By triangle inequality, either $P(|\operatorname{Re}(\xi_{ij})| \geq 1/2) > 0$ or $P(|\operatorname{Im}(\xi_{ij})| \geq 1/2) > 0$. Suppose that $P(|\operatorname{Re}(\xi_{ij})| \geq 1/2) > 0$. We may assume $P(\operatorname{Re}(\xi_{ij}) \geq 1/2) = p > 0$ by possibly considering the matrix $-W_n$. Define $V = \{z \in \mathbb{C} : \operatorname{Re}(z) > 1/2\}$. Considering the projections to the real line, we see that if $z_1, \ldots, z_n \in V$, then $|\sum_{j=1}^{n} z_j| \geq n/2$. Since V is symmetric with respect to the real axis, we have that $\xi_{ij} \in V$ if and only if $\xi_{ji} \in V$.

For any $n > 1$, we have that $P(\xi_{ij} \in V, 1 \leq i < j \leq n) = p^{n(n-1)/2} > 0$. Additionally, by unit variance hypothesis, we know that $P(|\xi_{ii}| < 2, 1 \leq i \leq n) = q^n > 0$. When both of these events happen, consider the vector $v = e_1 + e_2 + \ldots + e_n$ for large enough n:

$$\begin{aligned}
\|\frac{1}{\sqrt{n}}W_n v\| &= \frac{1}{\sqrt{n}}\left\|\sum_{i=1}^{n}(\sum_{j=1}^{n}\xi_{ij})e_i\right\| = \frac{1}{\sqrt{n}}\sqrt{\sum_{i=1}^{n}|\sum_{j=1}^{n}\xi_{ij}|^2} \\
&\geq \frac{1}{\sqrt{n}}\sqrt{\sum_{i=1}^{n}\left(|\sum_{j=1,j\neq i}^{n}\xi_{ij}|-|\xi_{ii}|\right)^2} \geq \frac{1}{\sqrt{n}}\sqrt{\sum_{i=1}^{n}\left(|\sum_{j=1,j\neq i}^{n}\xi_{ij}|-2\right)^2} \\
&\geq \frac{1}{\sqrt{n}}\sqrt{\sum_{i=1}^{n}\left(\frac{n-1}{2}-2\right)^2} \geq \sqrt{(cn)^2} = cn.
\end{aligned}$$

Here $c > 0$ is some constant. Now $\|v\| = \sqrt{n}$, so we know that the operator norm of $\frac{1}{\sqrt{n}}W_n$ is at least $c\sqrt{n}$. For large enough n this is always bigger than K. This means that $\mu_{\frac{1}{\sqrt{n}}W_n}\{x \in \mathbb{R} : |x| > K\} > 0)) \geq 1/n > 0$ with probability at least $p^{n(n-1)/2}q^n$.

Now we consider the case $P(|\operatorname{Im}(\xi_{ij})| \geq 1/2) > 0$. We may assume that $P(\operatorname{Im}(\xi_{ij}) \geq 1/2) > 0$. Representing the half-plane $V = \{z \in \mathbb{C} : \operatorname{Im}(z) \geq 1/2\}$ as a countable union, we conclude that $P(\operatorname{Im}(\xi_{ij} \in [r, 2r])) = p > 0$ for some $r \geq 1/2$. Defining v as before, we have

$$\|\frac{1}{\sqrt{n}}W_n v\| = \frac{1}{\sqrt{n}}\sqrt{\sum_{i=1}^{n}|\sum_{j=1}^{n}\xi_{ij}|^2}.$$

We know that ξ_{ij} is purely real if $i = j$. Then we know

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sqrt{\sum_{i=1}^{n}|\sum_{j=1}^{n}\xi_{ij}|^2} &= \frac{1}{\sqrt{n}}\sqrt{\sum_{i=1}^{n}|\sum_{j=1,j\neq i}^{n}\operatorname{Im}(\xi_{ij})|^2 + |\sum_{j=1}^{n}\operatorname{Re}(\xi_{ij})|^2} \\
&\geq \frac{1}{\sqrt{n}}\sqrt{\sum_{i=1}^{n}|\sum_{j=1,j\neq i}^{n}\operatorname{Im}(\xi_{ij})|^2}
\end{aligned}$$

We know that $P(\operatorname{Im}(\xi_{ij} \in [r, 2r], 1 \leq i < j \leq n) = p^{n(n-1)/2} > 0$. In the case where this happens we can estimate the inner sums as follows:

$$\begin{aligned}
|\sum_{j=1,j\neq i}^{n}\operatorname{Im}(\xi_{ij})| &\geq \sum_{j=1,j\neq i}^{n}\operatorname{Im}(\xi_{ij}) = \sum_{j=1}^{i-1}Im(\xi_{ij}) + \sum_{j=i+1}^{n}\operatorname{Im}(\xi_{ij}) \geq \sum_{j=1}^{i-1}r - \sum_{j=i+1}^{n}2r \\
&= (i-1)r - 2(n-i)r = (3i-(2n+1))r
\end{aligned}$$

We observe that this estimate is bigger that cnr for at least $c'n$ different i for some small absolute constants c and c'. This means we can estimate for large enough n

$$\frac{1}{\sqrt{n}}\sqrt{\sum_{i=1}^{n}|\sum_{j=1,j\neq i}^{n}\mathrm{Im}(\xi_{ij})|^2} \geq \frac{1}{\sqrt{n}}\sqrt{c'n(cnr)^2} = rn\sqrt{c'c^2} \geq Cn$$

for some constant C. Now once again, we know that the operator norm must be at least $C\sqrt{n}$, so the claim follows for large enough n. □

The previous lemma means that for a Wigner matrix W_n, the expected proportion of the eigenvalues λ with $|\lambda| > K\sqrt{n}$ becomes positive after n goes large enough. By semicircular law, this expectation goes to zero as n goes to infinity for $K > 2$.

Chapter 4

Operator norm of Wigner matrices

Having proved the semicircular law, we turn our attention to the operator norm of the Wigner matrices. We are going to prove that almost surely

$$\lim_{n\to\infty} \frac{\|W_n\|}{\sqrt{n}} = 2,$$

at least under our assumptions [13]. We may immediately prove a lower bound for the operator norm of Wigner matrices using the semicircular law because the semicircular distribution has a positive density near -2 and 2.

Theorem 4.1. (Bai-Yin theorem, lower bound.) *For an ensemble of Wigner matrices $(W_n)_{n\in\mathbb{Z}_+}$, we have for any $\varepsilon > 0$ almost surely*

$$\|W_n\| > (2-\varepsilon)\sqrt{n}$$

for large enough n. Here we denote the operator norm of matrix A with $\|A\|$. In fact, we have almost surely

$$\liminf_{n\to\infty} \|W_n\|/\sqrt{n} \geq 2.$$

Proof. By semicircular law, we know that the empirical spectral distribution of $(1/\sqrt{n})W_n$ converges to the semicircular distribution almost surely. Fix $\varepsilon > 0$. Let $g : \mathbb{R} \to [0,1]$ be a continuous function with $g(x) = 1$ if $|x| < 2-\varepsilon$ and $g(x) = 0$ if $|x| > 2-\varepsilon/2$. If $\|W_n\| < (2-\varepsilon)\sqrt{n}$, then

$$\int_{\mathbb{R}} g(x) d\mu_{\frac{1}{\sqrt{n}}W_n}(x) = 1.$$

But we also see that

$$\int_{\mathbb{R}} g(x) d\mu_{sc}(x) < 1.$$

This means that we can almost surely find N such that

$$\|W_n\| > (2-\varepsilon)\sqrt{n}$$

for any $n > N$. $\square$

We try to prove the associated upper bound next. While the semicircular law does in fact hold even without bounded fourth moments [12], the upper bound on the operator norm depends heavily on the assumption of bounded fourth moments, as illustrated by the following example.

Example 4.2. Let M_n be a real symmetric $n \times n$ random matrix with all independent identically distributed upper triangular coefficients having probability density function

$$f(x) = C\frac{\chi_{\{|x|>r\}}(x)}{|x|^5},$$

where the constants r and C are chosen so that this defines a probability density function of a random variable with zero mean and unit variance. Then the coefficients of M_n satisfy the weak estimate

$$P(|\xi_{ij}| > \lambda) \leq \frac{c}{\lambda^4},$$

but there is no constant s such that almost surely

$$\limsup_{n\to\infty} \|M_n\|/\sqrt{n} \leq s.$$

Namely, for any element ξ_{ij} of the matrix M_n we have (at least for large enough n)

$$P(|\xi_{ij}| \leq s\sqrt{n}) = 1 - \frac{C}{s^4n^2}.$$

As the upper triangular elements of the matrix are independent, we have

$$P(\|M_n\| \leq s\sqrt{n}) \leq P(|\xi_{ij}| \leq s\sqrt{n})^{\frac{n(n+1)}{2}} \leq \sqrt{\left(1 - \frac{C}{s^4n^2}\right)^{n^2}} \to e^{-C/2s^4} < 1.$$

This implies that the operator norm does not even converge in probability.

Nonetheless, as we assumed that the upper triangular coefficients have a finite fourth moment, we will see that the lower bound implied by the semicircular law is strict.

Theorem 4.3. (Bai-Yin theorem, upper bound.) *For an ensemble of Wigner matrices* $(W_n)_{n\in\mathbb{Z}_+}$*, we have almost surely*

$$\limsup_{n\to\infty} \|W_n\|/\sqrt{n} \le 2.$$

Combined with Theorem 4.1, this implies that almost surely

$$\lim_{n\to\infty} \|W_n\|/\sqrt{n} = 2.$$

To prove Theorem 4.3 we will adopt the approach used by Tao [10], splitting W_n into different parts and estimating them separately. We are also going to use *lacunary* sequences to obtain almost sure convergence. The sequences we consider are of form $n_m = \lfloor (1+\delta)^m \rfloor$ for some $\delta > 0$.

First, we estimate the diagonal elements. Letting D_n be the matrix of the diagonal elements, we have the following theorem.

Theorem 4.4. *Let X be a real random variable with zero mean and finite variance. Let $(D_n)_{n=1}^{\infty}$ be a random matrix ensemble with D_n being $n \times n$ diagonal matrix, all diagonal entries being independent identically distributed copies of X. Then almost surely*

$$\limsup_{n\to\infty} \|D_n\|/\sqrt{n} = 0.$$

Proof. Fix $\varepsilon > 0$. We are going to show the claim for powers of two, i.e. that almost surely

$$\limsup_{n\to\infty} \|D_{2^n}\|/\sqrt{2^n} \le \varepsilon.$$

This is sufficient because $\|D_{n-1}\| \le \|D_n\|$, so

$$\limsup_{n\to\infty} \|D_n\|/\sqrt{n} \le \limsup_{m\to\infty} \sqrt{2}\|D_{2^m}\|/\sqrt{2^m} \le \sqrt{2}\varepsilon.$$

We apply Borel-Cantelli lemma (Theorem A.1). We can use the geometric series formula to see that

$$\sum_{n=0}^{\infty} P(\|D_{2^n}\| \ge \sqrt{2^n}\varepsilon) \le \sum_{n=0}^{\infty} 2^n P(|X| \ge \sqrt{2^n}\varepsilon) \le \frac{2}{\varepsilon^2}\sum_{n=0}^{\infty} 2^n\varepsilon^2 P(\sqrt{2^n}\varepsilon \le |X| \le \sqrt{2^{n+1}}\varepsilon).$$

The above sum is the expectation of a random variable that is bounded above by $|X|^2$, so the sum is finite. We conclude that almost surely

$$\limsup_{n\to\infty} \|D_{2^n}\|/\sqrt{2^n} \le \varepsilon.$$

$\square$

Because $\|W_n\| \leq \|D_n\| + \|W_n - D_n\|$ by triangle inequality, we may assume that all the diagonal elements of W_n vanish. Next we do a lacunary reduction similar to the one used in the proof of Theorem 4.4. Let $\delta > 0$ be fixed. We define the sequence $(n_m)_{m=1}^{\infty}$ by setting $n_m = \lfloor (1+\delta)^m \rfloor$. Here $\lfloor x \rfloor$ is the largest integer at most x, the floor function at x. To see why this is useful, suppose that we proved an upper bound for the limit for all such sequences so that

$$\limsup_{m\to\infty} \|W_{n_m}\|/\sqrt{n_m} \leq a$$

for some absolute constant a independent from the chosen δ. Then as $\|W_n\| \leq \|W_{n+1}\|$, we have for any fixed δ almost surely

$$\limsup_{n\to\infty} \|W_n\|/\sqrt{n} \leq \limsup_{m\to\infty} \sqrt{1+\delta}\|W_{n_m}\|/\sqrt{n_m} \leq \sqrt{1+\delta}a.$$

As δ is arbitrary, it follows that $\limsup_{n\to\infty} \|W_n\|/\sqrt{n} \leq a$ almost surely.

Next, we are going to split the matrix W_n into two parts: a "small" part and a "big" part. Let $\eta > 0$ be sufficiently small: exact value of η will be fixed later. For the upper triangular entries ξ_{ij} of W_n, define $g_{ij} = \xi_{ij}\chi_{|\xi_{ij}|\leq n^{1/2-\eta}}$ and $b_{ij} = \xi_{ij} - g_{ij}$. We can then split the matrix $W_n = G_n + B_n$. For the matrix G_n the estimates are easier to do if we assume the off-diagonal elements to have zero mean. To justify this additional assumption, we consider a matrix $EG_n = (Eg_{ij})_{i,j=1}^{n}$. As $Eg_{ij} = -Eb_{ij}$, we see that $|Eg_{ij}| = |Eb_{ij}| \leq E|b_{ij}| \leq n^{-3/2+3\eta}E|b_{ij}|^4 \leq K/n^{3/2-3\eta}$. We fix $x \in \mathbb{C}^n$ with $\|x\| = 1$ and use the Cauchy-Schwartz inequality.

$$\begin{aligned}
\|EG_n x\| = \left(\sum_{i=1}^{n}\left|\sum_{j=1}^{n} Eg_{ij}x_j\right|^2\right)^{1/2} &\leq \left(\sum_{i=1}^{n}\sum_{j=1}^{n} n|Eg_{ij}x_j|^2\right)^{1/2} \\
\leq \left(\sum_{i=1}^{n} K^2 n^{-2+6\eta}\sum_{j=1}^{n}|x_j|^2\right)^{1/2} &= \left(K^2 n^{-1+6\eta}\right)^{1/2} = Kn^{-1/2+3\eta}
\end{aligned}$$

We see that the operator norm of EG_n is at most $Kn^{-1/2+3\eta}$, so in particular, if $\eta < 1/6$, it tends to zero surely as $n \to \infty$. This means that we can consider the hermitian matrix $H_n := G_n - EG_n$, with all diagonal elements equal to 0 and all the off-diagonal elements having zero mean and are surely bounded in absolute values by $2n^{1/2-\eta}$ for large enough values of n. Observe that all the off-diagonal elements of H_n have variance at most one, because $Var(h_{ij}) = E|g_{ij}|^2 - |Eg_{ij}|^2 \leq E|g_{ij}|^2 \leq E|\xi_{ij}|^2 = 1$.

By spectral theorem, H_n has n real eigenvalues. This implies that for a positive even integer k the matrix H_n^k has n non-negative eigenvalues. As the operator norm of the matrix is the maximum of the absolute value of the eigenvalues, we see that

$$\|H_n\|^k = \|H_n^k\| \leq tr(H_n^k).$$

We want to show that for some sequence of even integers $k = k(n)$ we have for any $\varepsilon > 0$ something like

$$Etr(H_n^k) \leq (2+\varepsilon)^k n^{k/2}$$

for large enough n. Allowing k to depend on n allows us to discard some extra factors of n at the cost of requiring more care when estimating the expected value.

We are therefore going to estimate the expected value of $tr(H_n^k)$. We see that

$$tr(H_n^k) = \sum_{i_0,i_1,\dots,i_{k-1}\in\{1,\dots,n\},i_k=i_0} \prod_{j=0}^{k-1} h_{i_j i_{j+1}}.$$

This sum can be understood to be taken over all tours of length k in a complete graph with n vertices. We also keep in mind that all off-diagonal elements h_{ij} have zero mean, at most unit variance and are surely bounded in absolute value by $2n^{1/2-\eta}$.

All the upper triangular elements are independent, and any lower triangular element h_{ij} is determined by the corresponding upper triangular element h_{ji}. As a result, if one edge $\{i_j, i_{j+1}\}$ appears in a tour only once, the expectation of that term vanishes. The only remaining terms have every edge appearing at least twice. For any tour of length k in the sum, let d be the number of different edges that appear in it, and let the lth edge appear $\alpha_l \geq 2$ times. Using triangle inequality for the remaining terms, we see that the expectation is at most

$$\begin{aligned} Etr(H_n^k) &\leq \sum_{d=1}^{k/2} \sum_{T(d,k,n)} \prod_{l=1}^{d} E|h_{ij}|^{\alpha_l} \leq \sum_{d=1}^{k/2} \sum_{T(d,k,n)} \prod_{l=1}^{d} (2n^{1/2-\eta})^{\alpha_l-2} E|h_{ij}|^2 \\ &= \sum_{d=1}^{k/2} \left(2n^{1/2-\eta}\right)^{k-2d} \#T(d,k,n). \end{aligned}$$

We used the fact that α_l sum up to k. The notation $\sum_{T(d,k,n)}$ means that the sum is taken over all tours of length k in a complete graph of n vertices that contain exactly d edges where each edge is traversed at least twice. The set of these tours is denoted by $T(d,k,n)$. We now need to estimate the number of such tours $\#T(d,k,n)$.

Lemma 4.5. *Let $k = 2\lfloor \log(n)^t \rfloor$ for a fixed exponent $t > 1$. Then for large enough n we have the estimate*

$$\#T(d,k,n) \leq 2^k n^{d+1} k^{k-2d} d^{k-2d+1}.$$

Proof. For any tour in $T(d, k, n)$, we will record sufficient information to reconstruct it. A tour is uniquely determined if we know the starting vertex and for every step which edge to take. An extremely naive approach is to record the starting vertex, and record for every one of k steps which one of the d edges we took. Additionally, for every edge, record the second endpoint of the edge for the first time we take it. This would give an upper bound $d^k n^{d+1}$, which is not sufficient. We need to be more careful.

Let v be the total number of vertices we visit. For every edge that takes us back to the vertex we have already visited, there are only $v - 1$ possible destinations. Another idea to reduce some unnecessary information is noting that when we leave a vertex for the last time, there is only one possible destination if we know the number of times every edge is traversed.

For every edge we visit, we make a list of d integers: the ith integer is the number of times we use the ith edge during the tour. As every edge has to be travelled at least twice and the integers sum to k, we can assign the remaining $k - 2d$ traversals to possibly any edge on the list. The number of ways to assign a possibly valid list is therefore at most d^{k-2d}.

We record the number of distinct vertices in the tour. The number v is between 2 and $d + 1$. We also make a list of the steps we first visit each vertex, which has at most $\binom{k}{v-1} \leq 2^k$ different possibilities as the first vertex is in the very beginning. We will also list the vertices we use, which has at most n^v possibilities.

We see that we have $d - v + 1$ edges that do not take us to any new vertex. For them, we record their first position in the tour, which has at most $\binom{k}{d-v+1} \leq k^{d-v+1}$ possibilities. We will also make a list of the vertices we arrive to in the first appearance of such an edge, and this list can be made in at most $(v - 1)^{d-v+1}$ different ways as we cannot go from a vertex to itself.

Say that a step is ambiguous if it is not the first time we take an edge and there are at least two edges from the current vertex that we have taken at least once and that we have not taken for the last time. Since we know for every edge the number of times we take it, every ambiguous step can be recognized with the information we have already recorded. For every ambiguous step, we record the edge we take. There are at most $v - 1$ possible choices for any such step.

How many ambiguous steps can there be? For any vertex apart from the starting one, we will at some point leave it for the final time, and this will not be an ambiguous step. Additionally, when we traverse any edge for the first time, the step is not ambiguous. Such traversal will not be the final outgoing traversal from the incoming vertex, since every edge is traversed at least twice. Therefore there are at most $k - d - v + 1$ ambiguous steps.

Let us clarify by example how to reconstruct the tour. Set $v = 4$, $k = 12$ and $d = 5$, and n large enough. We try to reconstruct the tour

$$(1, 2, 3, 1, 4, 2, 1, 3, 1, 4, 2, 3, 1).$$

The edges first appear in the order $(1, 2)$, $(2, 3)$, $(3, 1)$, $(1, 4)$ and $(4, 2)$. They are traversed 2, 2, 4, 2 and 2 times respectively. The steps we first visit every vertex apart from 1 are 1st step for vertex 2, 2nd step for vertex 3 and 4th step for the vertex 4. For the edges that do not take us to any new vertices, $(3, 1)$ appears for the first time in the 3rd step and $(4, 2)$ appears in the 5th step, having endpoints 1 and 2 respectively.

The 6th step is ambiguous, since we arrive in the vertex 2 and we have 3 possible choices for our step. We record that the 6th step should end in vertex 1. Similarly, steps 7, 8 and 9 are also ambiguous, and we record that the destination vertices are 3, 1 and 4 respectively.

We can reconstruct the tour from this data: start from 1, then the first appearances of the edges forces us to travel $(1, 2, 3, 1, 4, 2)$. Now we have some ambiguous steps, so we follow the directions as $(2, 1, 3, 1, 4)$. For the rest of the tour, we have only one possible edge to take, so the tour ends as $(4, 2, 3, 1)$.

We see that for any fixed v we have at most

$$d^{k-2d}2^k n^v k^{d-v+1}(v-1)^{d-v+1}(v-1)^{k-d-v+1} = 2^k d^{k-2d} n^v k^{d-v+1}(v-1)^{k-2v+2}$$

possible choices. We can trivially estimate $v - 1 \leq k$ and $d \leq k$. Additionally we have $k^3 \leq n$ for sufficiently large n. We get an upper bound

$$2^k d^{k-2d} n^v k^{d-v+1} k^{k-2d+2d-2v+2} \leq 2^k d^{k-2d} n^v n^{d-v+1} k^{k-2d} = 2^k d^{k-2d} n^{d+1} k^{k-2d}.$$

As there are d possible choices for v the claim follows. $\square$

Using the previous lemma, we have

$$\begin{aligned} Etr(H_n^k) &\leq \sum_{d=1}^{k/2} \left(2n^{1/2-\eta}\right)^{k-2d} \#T(d, k, n) \leq \sum_{d=1}^{k/2} \left(2n^{1/2-\eta}\right)^{k-2d} 2^k n^{d+1} k^{k-2d} d^{k-2d+1} \\ &\leq \sum_{d=1}^{k/2} \left(2n^{1/2-\eta}\right)^{k-2d} 2^k n^{d+1} k^{k-2d} k^{k-2d+1} = \sum_{d=1}^{k/2} 2^k n^{k/2+1} \left(2n^{-\eta}k^2\right)^{k-2d} k \\ &\leq \sum_{d=1}^{k/2} 2^k n^{k/2+1} k \leq 2^k n^{k/2+1} k^2. \end{aligned}$$

Here we used the fact that $\log(n)^t$ grows slower than any power function. Now we can finally find an upper bound for the operator norm of H_{n_m} for lacunary sequence $n_m = \lfloor (1+\delta)^m \rfloor$.

Theorem 4.6. *For the sequence* $(n_m)_{m=1}^{\infty}$ *we have almost surely*

$$\limsup_{m\to\infty} \|H_{n_m}\|/\sqrt{n_m} \leq 2.$$

Proof. Fix $\varepsilon > 0$. Let $k = 2\lfloor \log(n_m)^t \rfloor$. Then the preceding calculations show that for large enough m we have

$$E\|H_{n_m}\|^k \leq 2^k n_m^{k/2+1} k^2.$$

Using Markov's inequality tells us

$$P(\|H_{n_m}\| > (2+\varepsilon)\sqrt{n_m}) \leq \frac{2^k n_m^{k/2+1} k^2}{(2+\varepsilon)^k n_m^{k/2}} = \left(\frac{2}{2+\varepsilon}\right)^k n_m k^2.$$

If we denote $c = 2/(2+\varepsilon)$, then we see that the upper bound of the probability is of order

$$c^{\log(n_m)^t} e^{\log(n_m)} \log(n_m)^{2t} \leq \frac{1}{n_m^2}$$

for large enough m. This means that the series

$$\sum_m P(\|H_{n_m}\| > (2+\varepsilon)\sqrt{n_m})$$

converges and now by Borel-Cantelli lemma (A.1) almost surely

$$\limsup_{m\to\infty} \|H_{n_m}\|/\sqrt{n_m} \leq 2+\varepsilon.$$

As ε was arbitrary, we obtain

$$\limsup_{m\to\infty} \|H_{n_m}\|/\sqrt{n_m} \leq 2.$$

as wanted. □

Now we only need to estimate the norm of B_{n_m}.

Theorem 4.7. *For the sequence* $(n_m)_{m=1}^{\infty}$ *we have almost surely*

$$\limsup_{m\to\infty} \|B_{n_m}\|/\sqrt{n_m} \to 0.$$

Proof. Fix $\varepsilon > 0$. We will first prove that the matrix B_{n_m} has almost surely no entries larger than $\varepsilon\sqrt{n_m}$ as $m \to \infty$. We will estimate this probability, wanting to apply Borel-Cantelli lemma A.1. Estimating the infinite sum, we obtain

$$\begin{aligned}\sum_m P(\max|\xi_{ij}| > \varepsilon\sqrt{n_m}) &\leq \sum_m n_m^2 P(|\xi_{ij}| > \varepsilon\sqrt{n_m}) \\ &\leq \sum_m (1+\delta)^{2m} P(|\xi_{ij}| > \varepsilon\sqrt{(1+\delta)^{m-1}})\end{aligned}$$

We used the definition of $n_m = \lfloor(1+\delta)^m\rfloor$ in the final estimate, keeping in mind that we are only considering the indices m large enough so that $(1+\delta)^{m-1} \leq n_m$. Denote by E_m the event that $\varepsilon\sqrt{(1+\delta)^{m-1}} < |\xi_{ij}| \leq \varepsilon\sqrt{(1+\delta)^m}$. Then we can use the geometric series formula and the above estimate

$$\begin{aligned}\sum_m P(\max|\xi_{ij}| > \varepsilon\sqrt{n_m}) &= \sum_m \sum_{m' \geq m} (1+\delta)^{2m} P(E_{m'}) = \sum_{m'} \sum_{m \leq m'} (1+\delta)^{2m} P(E_{m'}) \\ &\leq \sum_{m'} \frac{(1+\delta)^{2m'+2} - 1}{(1+\delta)^2 - 1} P(E_{m'}) \\ &\leq \frac{(1+\delta)^2}{\varepsilon^4(2\delta + \delta^2)} \sum_{m'} (\varepsilon(1+\delta)^{m'/2})^4 P(E_{m'}).\end{aligned}$$

We may interpret the final sum as expectation of a random variable bounded above by $(1+\delta)^2|\xi_{ij}|^4$, and conclude that the sum is finite. Now Borel-Cantelli lemma applies.

We have shown that B_{n_m} has almost surely no elements larger that $\varepsilon\sqrt{n_m}$ as $m \to \infty$. If every row and every column happens to have at most one non-zero element, with all elements having absolute value less than $\sqrt{n_m}\varepsilon$, we could change the coordinates separately on both domain and range to conclude that $\|B_{n_m}\| \leq \varepsilon\sqrt{n_m}$. We will show that this happens almost surely as $m \to \infty$. We recall that the diagonal of B_{n_m} vanishes. We can estimate the probability that a given row or column has at least two non-zero elements using independence:

$$P \leq \frac{(n_m - 1)(n_m - 2)}{2} P(|\xi_{ij}| > n_m^{1/2-\eta})^2 \leq n_m^2 \left(\frac{K}{n_m^{2-4\eta}}\right)^2 = K^2 n_m^{8\eta-2}.$$

As there are n_m rows and n_m columns, we know that the probability that there is at least one row or column with at least two non-zero entries is at most

$$2K^2 n_m^{8\eta-1}.$$

As long as $\eta < 1/8$, we can use the geometric series formula to determine that

$$\sum_m 2K^2 n_m^{8\eta-1} \leq \sum_m 2K^2(1+\delta)^{(8\eta-1)m} < \infty.$$

Borel-Cantelli lemma tells that there is almost surely an index M such that for any $n_m > n_M$ all rows and all columns have at most one non-zero element, and the claim follows.

□

We have finally proved Theorem 4.3. As a summary, we split the matrix W_n into a diagonal part D_n, a small part G_n and a large part B_n, and obtained $\limsup_{n\to\infty} \|W_n\|/\sqrt{n} \leq 2$. We conclude with a short remark.

Remark 4.8. Assume that the coefficients of W_n are almost surely bounded. Going through our combinatorial lemma (Lemma 4.5), we observe that for any k there exists a constant C_k such that $Etr(W_n^k) = C_k n^{k/2+1} + o(n^{k/2+1})$, and the constant C_k does not depend on the distribution of the coefficients. This allows us to compute the moments of the empirical spectral distributions, and can be used to deduce the semicircular law. This approach is close to the original considerations of Wigner [4].

Appendix A

Some theorems from probability theory

Let (Ω, Γ, P) be a probability space. Let X_n be random variables for $n \in \mathbb{N}$, and X be another random variable. We say that X_n converges to X almost surely if for almost every $\omega \in \Omega$

$$\lim_{n \to \infty} X_n(\omega) = X(\omega).$$

Another way to state this in terms of probabilities of events is to require that for any $\varepsilon > 0$

$$P(|X_n - X| > \varepsilon \text{ for infinitely many } n) = 0,$$

or equivalently,

$$P(\limsup_{n \to \infty} |X_n - X| > \varepsilon) = \lim_{n \to \infty} P(\bigcup_{k \geq n} \{|X_k - X| > \varepsilon\}) = 0.$$

We say that X_n converges to X in probability if for any $\varepsilon > 0$

$$\lim_{n \to \infty} P(|X_n - X| > \varepsilon) = 0.$$

The convergence in probability is a weaker notion than the almost sure convergence: almost surely convergent random variables converge in probability, but the converse does not hold.

Theorem A.1. *Borel-Cantelli lemma, strong version. Suppose that we have sequences of random variables (X_n) and (Y_n). Suppose also that Y_n converges to a random variable Y almost surely, and that $\sum_n P(|X_n - Y_n| > \varepsilon) < \infty$ for any $\varepsilon > 0$. Then X_n converges to Y almost surely.*

Proof. Fix $\varepsilon > 0$. By triangle inequality, we know that for any n we have

$$\{|X_n - Y| > \varepsilon\} \subset \{|X_n - Y_n| > \varepsilon/2\} \cup \{|Y_n - Y| > \varepsilon/2\}.$$

Using this with union bound, we conclude that

$$P(\bigcup_{k\geq n}\{|X_k - Y| > \varepsilon\}) \leq P(\bigcup_{k\geq n}\{|X_k - Y_k| > \varepsilon/2\}) + P(\bigcup_{k\geq n}\{|Y_k - Y| > \varepsilon/2\}).$$

As Y_n converges to Y almost surely, we already know that

$$\lim_{n\to\infty} P(\bigcup_{k\geq n}\{|Y_k - Y| > \varepsilon/2\}) = 0.$$

Using sub-additivity, we obtain

$$P(\bigcup_{k\geq n}\{|X_k - Y_k| > \varepsilon/2\}) \leq \sum_{k\geq n} P(\{|X_k - Y_k| > \varepsilon/2\}),$$

and the right hand side tends to zero as n grows to infinity. This proves that X_n converges to Y almost surely. □

Borel-Cantelli lemma is usually stated for $Y_n = Y$ for some fixed random variable Y.

Theorem A.2. *Hoeffiding's lemma [6]. Let X be an almost surely bounded real random variable; in other words $X \in [a, b]$ almost surely. Then for any $t \geq 0$*

$$Ee^{tX} \leq e^{tEX} \exp(\frac{(b-a)^2}{8} t^2).$$

Proof. We may assume that $EX = 0$ and $b - a = 1$. Fix $t \geq 0$. An application of Jensen's inequality shows that for any $x \in [a, b]$

$$e^{tx} \leq (b - x)e^{ta} + (x - a)e^{tb}.$$

This implies that

$$\begin{aligned} Ee^{tX} &\leq (b - EX)e^{ta} + (EX - a)e^{tb} \\ &= e^{ta}(b - ae^t) \\ &= e^{ta+\ln(1+a-ae^t)} \\ &= e^{g(t)}. \end{aligned}$$

It remains to show that $g(t) \leq \frac{1}{8}t^2$. We note that $g(0) = g'(0) = 0$,

$$\begin{aligned} g'(t) &= a + \frac{-ae^t}{1+a-ae^t} \qquad \text{and} \\ g''(t) &= -\left(\frac{-ae^t}{1+a-ae^t}\right)^2 + \frac{-ae^t}{1+a-ae^t} \leq \frac{1}{4}. \end{aligned}$$

Hence by Taylor's theorem, $g(t) = \frac{g''(\xi_t)}{2}t^2 \leq \frac{1}{8}t^2$, proving the claim. $\square$

Appendix B

Hermitian matrices

An $n \times n$ complex matrix A is a Hermitian matrix if $a_{ij} = \overline{a_{ji}}$ for any $1 \leq i, j \leq n$. This is equivalent to having $\langle v, Aw \rangle = \langle Av, w \rangle$, in other words, A is a self-adjoint operator of the Hilbert space $\mathbb{C}^n$.

Theorem B.1. *Spectral theorem. A hermitian $n \times n$ matrix has n orthogonal eigenvectors, and their associated eigenvalues are real.*

Proof. We will prove by induction that any self-adjoint operator of n-dimensional Hilbert space has n eigenvectors with associated real eigenvalues. The claim holds trivially in zero-dimensional spaces.

Assume now that the claim holds in $(n-1)$-dimensional Hilbert spaces, and let A be a self-adjoint operator in n-dimensional Hilbert space. Consider the characteristic polynomial $\det(A - \lambda I_n)$. This is a polynomial of degree $n \geq 1$, so it has a complex root λ_n by the fundamental theorem of algebra.

As $\det(A - \lambda_n I_n) = 0$, we can find a vector $v_n \neq 0$ such that $Av_n = \lambda_n v_n$. As we now have an eigenvector, it remains to show that λ_n is real and that we can use the induction hypothesis.

Using self-adjointness, we see that

$$\lambda_n \langle v_n, v_n \rangle \;=\; \langle \lambda_n v_n, v_n \rangle = \langle Av_n, v_n \rangle = \langle v_n, Av_n \rangle = \langle v_n, \lambda_n v_n \rangle = \overline{\lambda_n} \langle v_n, v_n \rangle.$$

This proves that λ_n is real. To use the induction hypothesis, consider a vector w with $\langle v_n, w \rangle = 0$. We see that

$$\langle v_n, Aw \rangle \;=\; \langle Av_n, w \rangle = \lambda_n \langle v_n, w \rangle = 0,$$

so A maps the orthogonal complement of the line $\{tv_n | t \in \mathbb{C}\}$ to itself. But this defines a self-adjoint operator of $(n-1)$-dimensional Hilbert space and we may use the induction hypothesis to find its orthogonal eigenvectors $v_1, \dots, v_{n-1}$ with associated real eigenvalues $\lambda_1, \dots, \lambda_{n-1}$. We have found n orthogonal eigenvectors with real eigenvalues, so the claim has been proven. $\square$

Theorem B.2. *Min-max theorem of Courant-Fischer-Weyl. For an $n \times n$ Hermitian matrix A_n with eigenvalues $\lambda_1(A_n) \leq \ldots \leq \lambda_n(A_n)$, we have for any $1 \leq k \leq n$*

$$\lambda_k(A_n) = \inf_V \{\sup_x \{\langle Ax, x\rangle : x \in V, |x| = 1\} : \ V \textit{ is } k\textit{-dimensional subspace of } \mathbb{C}^n\}$$

and

$$\lambda_k(A_n) = \sup_V \{\inf_x \{\langle Ax, x\rangle : x \in V\} : \ V \textit{ is } (n-k+1)\textit{-dimensional subspace of } \mathbb{C}^n\}$$

Proof. Let (v_j) be the orthonormal eigenvectors associated with the eigenvalues $(\lambda_j(A_n))$. Taking $V = span(v_1, \dots, v_k)$, we see that for any $w = \sum_{i=1}^k c_i v_i$ with $|w| = 1$ we have $\langle A_n w, w\rangle = \sum_{i=1}^n \langle \lambda_i(A_n) c_i v_i, c_i v_i\rangle = \sum_{i=1}^k \lambda_i(A_n)|c_i|^2 \leq \max_{i \leq k} \lambda_i(A_n) = \lambda_k(A_n)$. Also, $\langle A_n v_k, v_k\rangle = \lambda_k(A_n)$. This proves that $\sup\{\langle Ax, x\rangle : x \in V, |x| = 1\} = \lambda_k(A_n)$ for this subspace V. This means that

$$\lambda_k(A_n) \geq \inf\{\sup\{\langle Ax, x\rangle : x \in V, |x| = 1\} : \ V \text{ is } k\text{-dimensional subspace of } \mathbb{C}^n\}$$

Let us prove the other direction. For any k-dimensional subspace V, we see that $V \cap span(v_k, \dots, v_n) \neq \{0\}$, because otherwise we have $n+1$ independent vectors on $\mathbb{C}^n$: the k vectors from the basis of V and $n-k+1$ vectors $v_k, \dots, v_n$. Taking a vector $w = \sum_{i=k}^n c_i v_i$ with $|w| = 1$ from this intersection, we get that $\langle A_n w, w\rangle = \sum_{i=k}^n \lambda_i(A_n)|c_i|^2 \geq \min_{i \geq k} \lambda_i(A_n) = \lambda_k(A_n)$. This means that $\sup\{\langle Ax, x\rangle : x \in V, |x| = 1\} \geq \lambda_k(A_n)$ for V, and the the first equality follows.

The second equality follows from the first as $\lambda_k(A_n) = -\lambda_{n-k+1}(-A_n)$. $\square$

Theorem B.3. *Cauchy interlacing formula. Let A_n be a Hermitian $n \times n$ matrix for $n \geq 2$ and A_n^l be its minor obtained by removing lth row and column. Denote the eigenvalues of A_n as $\lambda_1(A_n) \leq \ldots \leq \lambda_n(A_n)$ and the eigenvalues of A_n^l as $\lambda_1(A_n^l) \leq \ldots \leq \lambda_{n-1}(A_n^l)$. Then for any $1 \leq k < n$ we have*

$$\lambda_k(A_n) \leq \lambda_k(A_n^l) \leq \lambda_{k+1}(A_n).$$

Proof. It suffices to prove one of the inequalities, as the other follows from the identity $\lambda_k(A_n) = -\lambda_{n-k+1}(-A_n)$. Define the subspace $U = \{v \in \mathbb{C}_n : v_l = 0\} \subset \mathbb{C}^n$, which we can identify with $\mathbb{C}^{n-1}$. Let π_U be the projection $\mathbb{C}^n \to \mathbb{C}^{n-1}$, and φ_U be the natural embedding $\mathbb{C}^{n-1} \to \mathbb{C}^n$. We see that $A_n^l = \pi_U \circ A_n \circ \varphi_U$. Let $(v_j)_{j=1}^{n-1}$ be the images of the orthonormal eigenvectors of A_n^l under the embedding φ_U. By the previous theorem, we have

$$\lambda_{k+1}(A_n) = \inf_{V \subset \mathbb{C}^n, \dim V = k+1} \sup_x \{\langle A_n x, x\rangle : x \in V, |x| = 1\}$$

and, as $\langle \pi_U x, \pi_U y\rangle = \langle x, y\rangle$ whenever $y \in U$, we also have

$$\lambda_k(A_n^l) = \langle A_n v_k, v_k\rangle.$$

Let V be any $k+1$-dimensional subspace of $\mathbb{C}^n$. We see that $V \cap span(v_k, \ldots, v_{n-1}) \neq \{0\}$, because otherwise we would have $n+1$ linearly independent vectors in $\mathbb{C}^n$: $k+1$ vectors of the basis of V and $n-k$ vectors $v_k, \ldots, v_{n-1}$. If we pick w from this intersection with $|w| = 1$, then

$$\sup\{\langle A_n x, x\rangle : x \in V, |x| = 1\} \quad \geq \quad \langle A_n w, w\rangle \geq \lambda_k(A_n^l).$$

Taking the infimum over subspaces V shows that $\lambda_{k+1}(A_n) \geq \lambda_k(A_n^l)$.

□

Appendix C

Formulas for the elements of matrix inverse

Theorem C.1. *Cramer's rule. Let A be $n \times n$ invertible square matrix, and for any $1 \leq i, j \leq n$, let M_{ij} be the determinant of the $(n-1) \times (n-1)$ matrix obtained from A by removing ith row and jth column. Then the elements of the inverse matrix A^{-1} are given by*

$$A^{-1}_{ij} = (-1)^{i+j} \frac{M_{ji}}{\det(A)}.$$

Proof. Define matrix B by

$$B_{ij} = (-1)^{i+j} M_{ji}.$$

Fix $1 \leq i \leq n$. Now we can compute the diagonal element $(AB)_{ii}$ as

$$(AB)_{ii} = \sum_{k=1}^{n} A_{ik} B_{ki} = \sum_{k=1}^{n} (-1)^{i+k} A_{ik} M_{ik} = \det(A),$$

where the final equality is obtained by expanding the determinant along the ith row. If we instead have $1 \leq j \leq n$, $j \neq i$, we see that

$$(AB)_{ij} = \sum_{k=1}^{n} (-1)^{j+k} A_{ik} M_{jk}.$$

Now we see that this is the determinant of the $n \times n$ matrix obtained by replacing the jth row of A by the ith row. But then the rows of the matrix are linearly dependent, so the determinant must be 0.

We have obtained that $AB = \det(A) I_n$, and as the inverse matrix is unique, we see that $A^{-1} = \frac{1}{\det(A)} B$. □

Theorem C.2. *A formula for an element in inverse matrix based on Schur's complement. Let A_n be $n \times n$ invertible square matrix for $n \geq 2$, and denote*

$$A_n = \begin{pmatrix} A_{n-1} & C \\ R & a_{nn} \end{pmatrix},$$

where A_{n-1} is $(n-1) \times (n-1)$ matrix, R is $1 \times (n-1)$ row matrix and C $(n-1) \times 1$ column matrix. If both A_n and A_{n-1} are invertible, then we have

$$[A_n^{-1}]_{nn} = \frac{1}{a_{nn} - RA_{n-1}^{-1}C}.$$

Proof. We will solve the linear equation $A_n x = e_n$: the element we seek will be exactly x_n. This equation can written in form

$$\begin{pmatrix} A_{n-1} & C \\ R & a_{nn} \end{pmatrix} \begin{pmatrix} y \\ x_n \end{pmatrix} = \begin{pmatrix} A_{n-1}y + Cx_n \\ Ry + a_{nn}x_n \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

From the equation $A_{n-1}y + Cx_n = 0$ we can solve $y = -x_n A_{n-1}^{-1}C$. Substituting this to the equation $Ry + a_{nn}x_n = 1$ allows us to solve $x_n = (a_{nn} - RA_{n-1}^{-1}C)^{-1}$.

□

Bibliography

[1] Persi Diaconis: Patterns in Eigenvalues: the 70th Josiah Williard Gibbs Lecture, Bull. Amer. Math. Soc. (New Ser.), 40 (2) (2003), p. 155-177.

[2] John Wishart: The Generalised Product Moment Distribution in Samples from a Normal Multivariate Population, Biometrika, 20A (1/2) (1928), p. 32-52.

[3] A. Lane, R. G. Thomas, E. P. Wigner: Giant Resonance Interpretation of the Nucleon-Nucleus Interaction, Phys. Rev., 98 (1955), p. 693-701.

[4] Eugene P. Wigner: Characteristic vectors of bordered matrices with infinite dimensions, Ann. Math., 62 (3) (1955), p. 548-564.

[5] J. von Neumann, H. H. Goldstine: Numerical inverting of matrices of high order, Bull. Amer. Math. Soc., 53 (11) (1947), p. 1021-1099.

[6] Wassily Hoeffding: Probability Inequalities for Sums of Bounded Random Variables, Journ. Amer. Stat. Ass., 58 (301) (1963), p. 13-30.

[7] Colin McDiarmid: On the Method of Bounded Differences, Surveys in Combinatorics, 141 (1989), p. 148-188.

[8] Joseph L. Doob: Regularity Properties of Certain Families of Chance Variables, Trans. Amer. Math. Soc., 47 (3) (1940), p. 455-486

[9] Kazuoki Azuma: Weighted sums of certain dependent random variables, Tohoku Math. J., 19 (3) (1967), p. 357-367.

[10] Terence Tao: Topics in random matrix theory, American Mathematical Society, 2012.

[11] Zeév Rudnick, Peter Sarnak: Zeroes of principal L-functions and random matrix theory, Duke Math. J., 81 (2) (1996), p. 269-322.

[12] O. L. Arnold: On Wigner's semicircle law for the eigenvalues of random matrices, Z. Wahrsch. verw. Geb, 19 (1971), p. 191-198

[13] Z. D. Bai, Y. Q. Yin: Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix, Ann. Probab., 16 (4) (1988), p. 1729-1741